

BAFFI CAREFIN
Centre for Applied Research on International
Markets, Banking, Finance and Regulation

Working Paper

Katrien Antonio, Anastasios Bardoutsos, and Wilbert Ouburg

Bayesian Poisson log-bilinear models for
mortality projections with multiple populations

**BAFFI CAREFIN Centre Research Paper Series
No. 2015-5**

This Paper can be downloaded without charge from The Social
Science Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=2603569>

Bayesian Poisson log-bilinear models for mortality projections with multiple populations

Katrien Antonio^{1,2}, Anastasios Bardoutsos^{*1}, and Wilbert Ouburg^{2,3}

¹Faculty of Economics and Business, KU Leuven, Belgium.

²Faculty of Economics and Business, University of Amsterdam, The Netherlands.

³Delta Lloyd Groep, The Netherlands

April 30, 2015

Abstract

Life insurers, pension funds, health care providers and social security institutions face increasing expenses due to continuing improvements of mortality rates. The actuarial and demographic literature has introduced a myriad of (deterministic and stochastic) models to forecast mortality rates of single populations. This paper presents a Bayesian analysis of two related multi-population mortality models of log-bilinear type, designed for two or more populations. Using a larger set of data, multi-population mortality models allow joint modelling and projection of mortality rates by identifying characteristics shared by all sub-populations as well as sub-population specific effects on mortality. This is important when modeling and forecasting mortality of males and females, regions within a country and when dealing with index-based longevity hedges. Our first model is inspired by the two factor Lee & Carter model of [Renshaw and Haberman \(2003\)](#) and the common factor model of [Carter and Lee \(1992\)](#). The second model is the augmented common factor model of [Li and Lee \(2005\)](#). This paper approaches both models in a statistical way, using a Poisson distribution for the number of deaths at a certain age and in a certain time period. Moreover, we use Bayesian statistics to calibrate the models and to produce mortality forecasts. We develop the technicalities necessary for Markov Chain Monte Carlo ([MCMC]) simulations and provide software implementation (in R) for the models discussed in the paper. Key benefits of this approach are multiple. We jointly calibrate the Poisson likelihood for the number of deaths and the times series models imposed on the time dependent parameters, we enable full allowance for parameter uncertainty and we are able to handle missing data as well as small sample populations. We compare and contrast results from both models to the results obtained with a frequentist single population approach and a least squares estimation of the augmented common factor model.

Keywords: projected life tables, multi-population stochastic mortality models, Bayesian statistics, Poisson regression, one factor Lee & Carter model, two factor Lee & Carter model, Li & Lee model, augmented common factor model.

*Corresponding author. E-mail address: anastasios.bardoutsos@kuleuven.be

The authors gratefully acknowledge financial support from CAREFIN Bocconi Centre for Applied Research in Finance.

1 Introduction

The quantification of longevity risk in a systematic way requires stochastic mortality projection models, as [Barrieu et al. \(2012\)](#) indicate. The development of stochastic mortality models has received wide coverage in the actuarial, demographic and statistical literature. Starting from the seminal work by Lee & Carter ([LC]) ([Lee and Carter \(1992\)](#)), generalizations of their basic, log-bilinear, single (time dependent) factor model have been proposed. For example, [Renshaw and Haberman \(2003\)](#) propose a two factor version of the LC model and [Renshaw and Haberman \(2006\)](#) augment the one factor LC model with a cohort effect. [Cairns et al. \(2006\)](#) invent a stochastic two factor mortality model for pensioner age mortality, with focus on actuarial applications. Generalizations of the original Cairns, Blake & Dowd ([CBD]) model have been studied in [Cairns et al. \(2009\)](#) and [Mavros et al. \(2014\)](#). Combining ideas from the class of LC models on the one hand and the class of CBD models on the other hand, [Plat \(2009\)](#) and [Haberman and Renshaw \(2011\)](#) formulate multi-factor models which lead to an increased goodness-of-fit in sample. All these models focus on mortality data for a single country or population. They propose a structure for (a transformation of) the mortality or death rate at age x in period t using a combination of age, period and cohort effects. Mortality forecasts then follow from a two step approach. In a first step, unknown parameters are estimated, typically using Singular Value Decomposition ([SVD]) or Maximum Likelihood Estimation ([MLE]). For instance, [Brouhns et al. \(2002\)](#) introduce the (nowadays widely used) Poisson likelihood assumption for the number of deaths at a given age in a certain period. In a second step, the fitted values of time dependent parameters are forecasted using the AutoRegressive Integrated Moving Average ([ARIMA]) toolbox for time series modelling.

Recently the study of multi-population mortality models has received increasing attention within the literature on mortality forecasting. The strength of a multi-population model lies in the analysis of a bigger mortality data set, combining data from different sources (typically: genders, regions or countries), which allows more robust mortality modeling by identifying characteristics shared by all sub-populations as well as sub-population specific effects. The recent literature exposes multiple viewpoints within the framework of multi-population models. For example, [Carter and Lee \(1992\)](#) propose a common factor LC model where death rates of two populations (males and females) are jointly driven by a shared time-varying factor. [Li and Lee \(2005\)](#) analyse a collection of populations with similar socioeconomic characteristics using the augmented common factor model. This is a double log-bilinear mortality model which augments common age and period effects with sub-population specific age and period effects. [Cairns et al. \(2011\)](#) focus on the joint analysis of male and female mortality data, as well as the analysis of a reference population data set and portfolio specific mortality data, using an Age-Period-Cohort ([APC]) model with a stochastic spread around an averaged (for the male/female case) or a dominant (for the portfolio/population case) period effect. [Dowd et al. \(2011\)](#) introduce a gravity model for the analysis of two related but different sized populations. The dominant population is modelled using an APC model, whereas the subordinate population is expressed in terms of spread or deviation from the larger population. [Jarner and Kryger \(2011\)](#) use the SAINT model for joint modeling of Danish data with European-wide population data. [Wan et al. \(2013\)](#) model the dominant population, constructed by grouping the data of thirteen countries¹ into a reference population, using the Plat model (see [Plat \(2009\)](#)). The spread between the dominant population and the subordinate population, namely Switzerland, is modelled with a three factor LC

¹ Austria, Denmark, England-Wales, Finland, France, West Germany, Japan, Netherlands, Norway, Spain, Sweden, Switzerland and USA.

model. [Hatzopoulos and Haberman \(2013\)](#) introduce a methodology based on generalized linear models to analyse mortality dynamics for a pool of countries. [Danesi et al. \(2015\)](#) formulates and compares ten multi-population models using data from eighteen Italian regions. Apart from being studied in academic literature, the multi-population approach recently found its way to official census projections. For example in the Netherlands, the Central Bureau of Statistics ([CBS])² uses the [Li and Lee \(2005\)](#) ([LL]) model for Dutch data, combined with data from ten other European countries³, for its official 2012-2060 mortality projection (see [Stoeldraijer et al. \(2013a\)](#) and [Stoeldraijer et al. \(2013b\)](#)). The 2014 projection of the Royal Dutch Actuarial Association⁴ also uses the model of [Li and Lee \(2005\)](#) on a collection of European countries, as described in [Koninklijk Actuariel Genootschap \(2014\)](#).

The method used for calibrating parameters and projecting time dependent parameters is a crucial aspect in the process of constructing carefully designed mortality forecasts. Within the multi-population mortality literature, the original paper by [Li and Lee \(2005\)](#) uses a three step approach. In the first and second step, age and time dependent parameters are calibrated by applying twice the SVD technique. This is similar to the original [Lee and Carter \(1992\)](#) paper where SVD is used for a single population log-bilinear model. The first application of SVD estimates the common bilinear structure of age and period effects, and the second application of SVD extracts the population specific bilinear effects. In a third step, using appropriate time series models, the common period effect and the country specific period effect are projected forward. This method is used, for instance, by the Dutch Central Bureau of Statistics for their 2012-2060 projection model (see [Stoeldraijer et al. \(2013b\)](#)). As an alternative for SVD, the use of maximum likelihood estimation with Poisson likelihood is the widespread and preferred approach for calibration of single population mortality models, since the paper of [Brouhns et al. \(2002\)](#). Within the multi-population mortality literature, [Li \(2013\)](#) uses a conditional maximum likelihood approach to calibrate the LL model for Australian male and female data. [Li \(2013\)](#) first calibrates the common parameters of the LL model by maximizing the Poisson likelihood corresponding to a LC model on the total population. While fixing the common parameters to their estimates obtained in the first step, the population specific parameters are estimated by optimizing the Poisson likelihood of a second, country specific LC model.

The contributions of our paper are as follows. We formulate two multi-population models of LC type, inspired by [Carter and Lee \(1992\)](#), [Renshaw and Haberman \(2003\)](#) and [Li and Lee \(2005\)](#) and calibrate these models in a Bayesian way, while assuming a Poisson likelihood. To the best of our knowledge, this paper is the first attempt to calibrate multi-population mortality models of log-bilinear LC type in a Bayesian way. We illustrate the methodology on two empirical studies. The use of Bayesian statistics offers at least three advantages. Firstly, the calibration and forecast steps are combined, which leads to more consistent estimates of the period effects (see [Cairns et al. \(2011\)](#)). Secondly, the Bayesian approach provides a natural framework for incorporating parameter uncertainty (or: risk) in mortality forecasts, which is relevant - for example - in the new insurance regulatory framework of Solvency II. Lastly, the Bayesian approach allows adequate handling of small populations and missing data, which is relevant for instance when modelling portfolio data or data on regions of a country. Similar to [Czado et al. \(2005\)](#) for a single population LC model with Poisson likelihood, our paper covers all technical details of the MCMC steps necessary for the Bayesian calibration of the multi-population mortality models under study. Our paper therefore directly extends [Czado et al.](#)

²See www.cbs.nl.

³Denmark, England-Wales, Finland, France, Italy, Norway, Spain, Sweden, Switzerland, and West Germany.

⁴See www.ag-ai.nl.

(2005) to the multi-population setting.

This paper contributes to the literature on Bayesian estimation of mortality models. Within this literature [Bray \(2002\)](#) provides a discussion and comparison of Bayesian univariate APC models, used in medical statistics. [Pedroza \(2006\)](#) implements the Kalman Filter in the LC model assuming Gaussian likelihood. [Li et al. \(2015\)](#) use the Kalman Filter not only to calibrate the models of LC and CBD, but also to choose the optimal calibration period. [Kogure and Yoshiyuki \(2010\)](#) use a Bayesian LC model with Gaussian likelihood to price longevity risk. [Fushimi and Kogure \(2014\)](#) extend the original approach of [Kogure and Yoshiyuki \(2010\)](#) by using a two factor LC model for pricing under stochastic interest rate. [Reichmuth and Sarferaz \(2008\)](#) apply the Bayesian methodology to a [Renshaw and Haberman \(2006\)](#) type of mortality model. [Mavros et al. \(2014\)](#) present a new approach for the CBD model that is able to capture local interdependencies among ages and periods. They use Bayesian statistics to calibrate the model, while assuming Poisson likelihood for the number of deaths. Also in the field of demography the use of Bayesian statistics recently became more popular. [Abel et al. \(2010\)](#) consider a Bayesian autoregressive model with stochastic volatility to forecast population growth rates and point out the attractive properties of the Bayesian methodology. [Giroi and King \(2008\)](#) present a Bayesian framework for mortality forecasting with smoothing priors and covariate information such as income and smoking behavior. [Raftery et al. \(2013\)](#) propose a Bayesian hierarchical model producing probabilistic forecasts of period life expectancy at birth for all the countries of the world to 2100. In the field of parametric mortality laws [Dellaportas et al. \(2001\)](#) and [Njenga and Sherris \(2011\)](#) apply Bayesian estimation techniques to calibrate the Heligman and Pollard mortality law. With respect to Bayesian mortality models for multiple populations, [Cairns et al. \(2011\)](#) present a two-population APC model for the joint analysis of male and female mortality data, as well as the analysis of a reference population and portfolio specific mortality data. [Riebler and Held \(2010\)](#) and [Riebler et al. \(2012\)](#) focus on the Bayesian multivariate APC model with smoothing priors for age, period and cohort effects. More specifically, [Riebler and Held \(2010\)](#) assume no correlation between the population specific period effects, whereas [Riebler et al. \(2012\)](#) use a correlated, multivariate version of the same smoothing prior.

This paper is structured as follows: Section 2 introduces the multi-population mortality models of interest, discusses the likelihood assumptions and imposed parameter constraints. We formulate in Section 3 the prior assumptions for all the parameters and hyperparameters. Section 4 derives the MCMC steps leading to simulations from the posterior distribution of the parameters. Numerical illustrations are presented in Section 5. Section 6 concludes.

2 Poisson log-bilinear models for mortality projections with multiple populations

2.1 Literature overview

[Lee and Carter \(1992\)](#) introduce the pioneering model for mortality modelling and forecasting. They propose the following bilinear structure

$$\log \mu_{x,t} = \alpha_x + \beta_x \kappa_t, \quad (1)$$

where $\mu_{x,t}$ denotes the force of mortality for a person exactly aged x at time t . Hereby, $x \in \{x_{min}, \dots, x_{max}\}$ denote the ages available in the data, and $t \in \{t_{min}, \dots, t_{max}\}$ are the observed

periods. Under the assumption,

$$\mu_{x+\xi_1,t+\xi_2} = \mu_{x,t},$$

for $0 \leq \xi_1, \xi_2 < 1$, the maximum likelihood estimator for $\mu_{x,t}$ is the unsmoothed death rate $m_{x,t} := d_{x,t}/E_{x,t}$, where $d_{x,t}$ is the observed number of deaths aged x at time t and $E_{x,t}$ the corresponding observed exposure at risk. We refer to [Pitacco et al. \(2009\)](#) for more details. Assuming the following set of widely used identifiability constraints

$$\sum_x \beta_x = 1 \text{ and } \sum_t \kappa_t = 0, \quad (2)$$

an estimate for α_x is given by

$$\hat{\alpha}_x = \frac{1}{t_{max} - t_{min} + 1} \sum_{t=t_{min}}^{t_{max}} \log m_{x,t}. \quad (3)$$

For an overview of extensions and variations of the basic LC model in (1), designed for single populations, we refer to [Cairns et al. \(2009\)](#) and [Haberman and Renshaw \(2011\)](#). We specifically mention the extension of [Renshaw and Haberman \(2003\)](#), who propose a two factor, log-bilinear version of the LC model

$$\log \mu_{x,t} = \alpha_x + \beta_{x,1}\kappa_{t,1} + \beta_{x,2}\kappa_{t,2}, \quad (4)$$

which they calibrate using a Gaussian as well as a Poisson likelihood. This two factor model – say [LC-2] – avoids the trivial correlation structure of the original LC model (see [Cairns et al. \(2009\)](#) for a discussion) through the inclusion of an extra time dependent effect. Moreover, by including an extra bilinear component (i.e. $\beta_{x,2}\kappa_{t,2}$) the LC-2 model is able to capture age and time dynamics in a better way. This is confirmed by the work of [Hyndman and Ullah \(2007\)](#) who allow for multiple components in their functional Principal Components Analysis ([PCA]) approach for modelling smoothed log mortality rates. In contrast to the single principal component approach of [Lee and Carter \(1992\)](#), [Hyndman et al. \(2013\)](#) mention ‘*we use up to six (principal components) (Hyndman and Ullah (2007)) although in practice, only the first two have much effect for mortality*’ (see page 281).

In their search for a stochastic model incorporating dependence between the population underlying a hedging instrument and the population that is hedged, [Li and Hardy \(2011\)](#) consider four solutions, starting from the [Lee and Carter \(1992\)](#) model. With $\mu_{x,t}^{(i)}$ denoting the force of mortality at age x during year t for population $i = 1, 2$, the four scenarios are:

- (1) *independent* modeling of both populations by two unrelated LC models

$$\log \mu_{x,t}^{(i)} = \alpha_x^{(i)} + \beta_x^{(i)} \kappa_t^{(i)}. \quad (5)$$

Implementation of this model is straightforward, but dependence between the mortality rates of the populations is completely ignored;

- (2) the *joint* κ model, as proposed in [Carter and Lee \(1992\)](#)

$$\log \mu_{x,t}^{(i)} = \alpha_x^{(i)} + \beta_x^{(i)} \kappa_t, \quad (6)$$

with κ_t a common trend, driving the time dynamics of both populations. As [Li and Hardy \(2011\)](#) (page 183) mention, this is a parsimonious way of linking the two populations and avoiding them to evolve in completely different ways. However, it creates perfect correlation between the mortality rates of the two populations and does not allow for population specific time dynamics;

- (3) a *co-integrated* Lee–Carter model, i.e. the model in (5) where $\boldsymbol{\kappa}_t = (\kappa_t^{(1)}, \kappa_t^{(2)})'$ is modelled with a bivariate random walk with drift and the presence of co-integration is verified. In the Li and Hardy (2011) example of female data from Canada and the United States co-integration is detected;
- (4) the *augmented common* factor model of Li and Lee (2005) for two populations

$$\log \mu_{x,t}^{(i)} = \alpha_x^{(i)} + B_x K_t + \beta_x^{(i)} \kappa_t^{(i)}. \quad (7)$$

This model uses age and period effects shared by all populations in the sample (denoted B_x and K_t), next to population i 's specific deviation driven by $\beta_x^{(i)}$ and $\kappa_t^{(i)}$. We can rewrite equation (7) as follows:

$$\log \mu_{x,t}^{(i)} = \tilde{\alpha}_x^{(i)} + \beta_x^{(i)} \kappa_t^{(i)} + A_x + B_x K_t, \quad (8)$$

where $\tilde{\alpha}_x^{(i)} := \alpha_x^{(i)} - A_x$.

For the calibration of (7), Li and Lee (2005) (as well as Li and Hardy (2011) and Stoeldraijer et al. (2013a)) use a two step approach. Firstly, they estimate B_x and K_t by calibrating the LC model, using SVD, to the global, overall population created by aggregating the observed deaths and exposures over all populations in the study. This corresponds to the following minimization problem

$$\min_{B_x, K_t} \left(\log \mu_{x,t}^{(\bullet)} - A_x - B_x K_t \right)^2, \quad (9)$$

where $A_x = \frac{1}{t_{max} - t_{min} + 1} \sum_{t_{min}}^{t_{max}} \log \mu_{x,t}^{(\bullet)}$, $\mu_{x,t}^{(\bullet)} = \sum_{i=1}^p d_{x,t}^{(i)} / \sum_{i=1}^p E_{x,t}^{(i)}$ and p is the number of populations. Hereby, $d_{x,t}^{(i)}$ is the observed number of deaths for population i , aged x in period t and $E_{x,t}^{(i)}$ is the corresponding exposure measure. Secondly, SVD is used again for each population separately, namely on the matrix $z_{x,t}^{(i)} = \log \mu_{x,t}^{(i)} - \alpha_x^{(i)} - B_x K_t$, with $\alpha_x^{(i)}$ estimated as in (3). As such, the population specific parameters $\beta_x^{(i)}$ and $\kappa_t^{(i)}$ are estimated. Finally, the Li and Lee (2005) approach identifies time series models for K_t and $\kappa_t^{(i)}$ for $i = 1, 2, \dots, p$ and projects them independently. In the set-up of Li and Lee (2005), the product $B_x K_t$ models the long term trend and age specific fluctuations shared by all populations. Inspired by the LC model for a single population, an appropriate time series model for K_t is the random walk with drift ([RWD]). The product $\beta_x^{(i)} \kappa_t^{(i)}$ describes short term deviations from the common trend, specific to population i . Hereby $\kappa_t^{(i)}$ tends to a constant level in the long run. As a result, Li and Lee (2005) propose modeling $\kappa_t^{(i)}$ as a mean reverting process, for example AR(1).

2.2 Model formulations, likelihoods, parameter constraints and coherence

Model formulations and likelihood assumptions. The goal of this paper is the technical development and empirical illustration of Bayesian log-bilinear multi-population mortality models. We propose and study two specific multi-population mortality models which meet the following goals. Firstly, the models allow joint modeling of multiple (say p , where $p \geq 2$) populations by identifying a common as well as a population specific time trend. The common trend captures the downward, typically rather linear trend that would be identified with a LC model

on the whole group. Secondly, the models avoid the ad hoc, inefficient three step approach to mortality calibration and forecasting, as used by [Li and Lee \(2005\)](#). Instead, we properly define the random variables of interest and specify reasonable distributional assumptions, similar to [Brouhns et al. \(2002\)](#). We identify the time dependent effects as dynamic latent factors and specify an appropriate time series structure for them. Bayesian inference then allows a joint approach to calibration and forecasting. Thirdly, the models have good in sample performance, with enough flexibility to capture age and time effects.

For the first model the LC-2 specification from [Renshaw and Haberman \(2003\)](#) is the starting point. For example, [Fushimi and Kogure \(2014\)](#) use this LC-2 model when calibrating longevity hedges. Let $D_{x,t}^{(i)}$ be the random variable representing the number of deaths at age x in year t of population i , and $E_{x,t}^{(i)}$ is the corresponding exposure. We propose the following multi-population mortality model:

$$\begin{aligned} D_{x,t}^{(i)} | \mu_{x,t}^{(i)} &\sim \text{Poisson} \left(E_{x,t}^{(i)} \cdot \mu_{x,t}^{(i)} \right), \\ \text{with } \mu_{x,t}^{(i)} | K_t, \kappa_t^{(i)} &= \exp \left(\alpha_x^{(i)} + \beta_{x,1}^{(i)} K_t + \beta_{x,2}^{(i)} \kappa_t^{(i)} \right). \end{aligned} \quad (10)$$

K_t is a dynamic latent factor, shared by all populations, and $\kappa_t^{(i)}$ a population specific, dynamic latent factor. K_t and $\kappa_t^{(i)}$ are assumed mutually independent for each i . Moreover, we assume that, given K_t and $\kappa_t^{(i)}$, the $D_{x,t}^{(i)}$ random variables are mutually independent. The structure imposed on the logarithm of the force of mortality is the two factor LC-2 model of [Renshaw and Haberman \(2003\)](#), where the populations under study are linked by a common period effect, K_t , as in [Carter and Lee \(1992\)](#). Population specific time dynamics are governed by $\kappa_t^{(i)}$. We introduce the acronym LC-2,t for this particular model, where the subscript ‘2’ refers to the two factor structure and ‘t’ denotes the common period effect. The second model under study is the augmented common factor LL model. Similar to [Li \(2013\)](#) and [Koninklijk Actuarieel Genootschap \(2014\)](#) we study this model under the Poisson likelihood assumption, namely

$$\begin{aligned} D_{x,t}^{(i)} | \mu_{x,t}^{(i)} &\sim \text{Poisson} \left(E_{x,t}^{(i)} \cdot \mu_{x,t}^{(i)} \right), \\ \text{with } \mu_{x,t}^{(i)} | K_t, \kappa_t^{(i)} &= \exp \left(A_x + B_x K_t + \alpha_x^{(i)} + \beta_x^{(i)} \kappa_t^{(i)} \right). \end{aligned} \quad (11)$$

Again, we assume that the $D_{x,t}^{(i)}$ random variables, given K_t and $\kappa_t^{(i)}$, are mutually independent for all x, t and i combinations.

Parameter constraints. It is well known (see [Lee and Carter \(1992\)](#)) that the LC model requires constraints to identify the parameters used in the model. For the model in (1) the most commonly used constraints are $\sum_{t=t_{min}}^{t_{max}} \kappa_t = 0$ and $\sum_{x=x_{min}}^{x_{max}} \beta_x = 1$. We calibrate the new LC-2,t model, as specified in (10), while imposing the following set of $2 + 3 \times p$ constraints. As in the LC model we center the time dependent parameters, thus

$$\sum_{t=t_{min}}^{t_{max}} K_t = 0, \quad (12)$$

$$\sum_{t=t_{min}}^{t_{max}} \kappa_t^{(i)} = 0, \quad \text{for } i = 1, 2, \dots, p. \quad (13)$$

Since the $\beta_{x,1}^{(i)}$ and $\beta_{x,2}^{(i)}$ can only be identified up to a multiplicative constant, we impose

$$\frac{1}{p} \sum_{i=1}^p \sum_{x=x_{\min}}^{x_{\max}} \beta_{x,1}^{(i)} = 1, \quad (14)$$

$$\left\| \beta_{x,2}^{(i)} \right\|_2 = 1, \text{ for } i = 1, 2, \dots, p, \quad (15)$$

where $\|\cdot\|_2$ denotes the Euclidean norm. Both sets of constraints are slightly different from what is often used in the LC framework, see (2). We avoid using the constraint $\sum_x \beta_{x,1}^{(i)} = 1$ since its implementation requires multiplying the common K_t with a constant that is population i dependent. The alternative constraint proposed in (14) circumvents this issue. The second constraint is motivated by better numerical performance in our simulations. Indeed, the implementation of the constraint imposed on the $\beta_{x,2}^{(i)}$'s in (2) requires dividing by $\sum_{x=x_{\min}}^{x_{\max}} \beta_{x,2}^{(i)}$ and this creates numerical problems when this sum is close to 0.

As in factor analysis (see [Mardia et al. \(2008\)](#), Chapter 9) we want the latent factors K_t and $\kappa_t^{(i)}$ ($i = 1, \dots, p$) to be uncorrelated. Since they both have mean zero (due to (12) and (13)), we impose orthogonality of K_t and $\kappa_t^{(i)}$. We define

$$r_{\kappa}^{(i)} = \frac{\sum_t K_t \cdot \kappa_t^{(i)}}{\sum_t K_t^2}. \quad (16)$$

The vector $\tilde{\kappa}^{(i)} = \kappa^{(i)} - r_{\kappa}^{(i)} \mathbf{K}$ is then orthogonal to $\mathbf{K}$, with $\mathbf{K} = (K_{t_{\min}}, \dots, K_{t_{\max}})$ and $\kappa^{(i)} = (\kappa_{t_{\min}}^{(i)}, \dots, \kappa_{t_{\max}}^{(i)})$. The implementation of the orthogonality constraint is then straightforward and does not change the specification of $\mu_{x,t}^{(i)}$ (and therefore the likelihood assumption). Indeed,

$$\begin{aligned} \log \mu_{x,t}^{(i)} &= \alpha_x^{(i)} + \beta_{x,1}^{(i)} \cdot K_t + \beta_{x,2}^{(i)} \cdot \kappa_t^{(i)} \\ &= \alpha_x^{(i)} + \left(\beta_{x,1}^{(i)} + r_{\kappa}^{(i)} \beta_{x,2}^{(i)} \right) \cdot K_t + \beta_{x,2}^{(i)} \cdot \left(\kappa_t^{(i)} - r_{\kappa}^{(i)} K_t \right) \\ &= \alpha_x^{(i)} + \tilde{\beta}_{x,1}^{(i)} \cdot K_t + \beta_{x,2}^{(i)} \cdot \tilde{\kappa}_t^{(i)} = \log \tilde{\mu}_{x,t}^{(i)}. \end{aligned} \quad (17)$$

[Niu and Melenberg \(2014\)](#) apply a similar orthogonality constraint when analyzing a LC type of model with an observable time dependent covariate, next to the κ_t effect.

We apply the constraints in a specific order. Firstly, we apply (12) and (13). Then we implement the orthogonality constraint. Note that if we have latent factors K_t and $\kappa_t^{(i)}$ satisfying (12) and (13), then also the transformed $\tilde{\kappa}_t^{(i)}$ (using (16)) will satisfy (13). Finally, we apply the constraints in (14) and (15). In Section 4 we use this particular order when updating the parameters using Gibbs sampling.

In the augmented common factor model (11) we can not impose the orthogonality of the latent effects in a straightforward way, since B_x is not depending on (i) and can therefore not be manipulated with a constant depending on (i) . We circumvent this issue by calibrating the LL model under the assumption of Poisson likelihood in two steps. In the first step we fit a LC model to $D_{x,t}^{(\bullet)} := \sum_{i=1}^p D_{x,t}^{(i)}$ and $E_{x,t}^{(\bullet)} := \sum_{i=1}^p E_{x,t}^{(i)}$:

$$\begin{aligned} D_{x,t}^{(\bullet)} | \mu_{x,t} &\sim \text{Poisson} \left(E_{x,t}^{(\bullet)} \cdot \mu_{x,t} \right), \\ \text{with } \mu_{x,t} | K_t &= \exp(A_x + B_x K_t), \end{aligned} \quad (18)$$

under the constraints (12) and $\sum_{x=x_{min}}^{x_{max}} B_x = 1$. In the second step we fit a LC model for each population separately, namely

$$\begin{aligned} D_{x,t}^{(i)} | \mu_{x,t}^{(i)}, \mu_{x,t} &\sim \text{Poisson} \left(E_{x,t}^{(i)} \cdot \mu_{x,t} \cdot \mu_{x,t}^{(i)} \right), \\ \text{with } \mu_{x,t}^{(i)} | \kappa_t^{(i)} &= \exp \left(\alpha_x^{(i)} + \beta_x^{(i)} \kappa_t^{(i)} \right), \end{aligned} \quad (19)$$

while imposing the constraints (13) and $\|\beta_x^{(i)}\|_2 = 1$ for $i = 1, 2, \dots, p$. Li (2013) and Koninklijk Actuarieel Genootschap (2014) apply this strategy to calibrate the LL model in a frequentist way, assuming Poisson likelihood. Using conditional maximum likelihood these authors fix the common parameters obtained in the first step and then calibrate the LC model in the second step. When calibrating the LL model without the two step strategy, Li (2013) acknowledges (page 125) ‘to ensure convergence, further constraints may need to be imposed; for example, the parameters in the common factor and the additional factors are all orthogonal to one another.’ Regarding the two step (or: conditional maximum likelihood) approach, Li (2013) (page 126) mentions ‘This fitting strategy provides a convenient way of refining modelling. There are basically no convergence problems encountered during our estimation process.’

Coherence. When studying mortality models for multiple countries the notion of coherence is important. Li and Lee (2005) derive a necessary and sufficient condition to avoid long run divergence of mortality forecasts, namely the limit, $\lim_{t \rightarrow +\infty} (\log \mu_{x,t}^{(i)} - \log \mu_{x,t}^{(j)})$, will not diverge for $i \neq j$. Because of its specific model formulation (with a common B_x and K_t) and the AR(1) time series model for the country specific time effects, the LL model leads to coherent forecasts. Due to the inclusion of population specific dynamic effects, a population’s age specific death rate may differ from the others in the short run, but this effect will taper off in the long term. Hyndman et al. (2013) formalize the notion of ‘non-divergent’ or coherent forecasts as follows: ‘[coherence is] the convergence of the ratios of the forecast age-specific death rates from any two subpopulations to appropriate constants’ (see page 276). In this study of log-bilinear multi-population mortality models the LL model is a coherent model according to definition of Hyndman et al. (2013). On the other hand, the LC-2,t model is not (necessarily) a coherent model. Despite being a more flexible model with computational advantages (see the discussion on parameter constraints), the inclusion of the country specific parameter $\beta_{x,1}^{(i)}$ for $i = 1, 2, \dots, p$ may create divergence of the projections for countries i and j (with $i \neq j$) in the long run. In the empirical studies of Section 4 we illustrate the performance of both log-bilinear mortality models under study.

3 Prior distributions

Below we present the priors for the LC-2,t model. We only explicitly mention priors for the LL model when these are different from the LC-2,t specifications.

3.1 Prior distributions for the time parameters

In contrast to the frequentist approach, a remarkable feature of the Bayesian methodology is the integration of the calibration and projection steps, by specifying a time series prior for the

time dependent parameters. We choose the common, time dependent, latent effects K_t to follow an autoregressive process of order one (AR(1)) around a linear trend. As such we follow the trend stationary approach proposed in the Bayesian LC model of [Czado et al. \(2005\)](#), and we leave the difference stationary model for K_t for further research, see [Van Berkum et al. \(2014\)](#) for a discussion of both approaches. Following [Li and Lee \(2005\)](#) all population specific latent effects $\kappa_t^{(i)}$ follow an AR(1) process without intercept. This ensures that in the long run $\kappa_t^{(i)}$ converges to zero, which implies that the country specific time trends fade out. K_t and $\kappa_t^{(i)}$ are independent and so are $\kappa_t^{(i)}$ and $\kappa_t^{(j)}$ for $i \neq j$, see the original paper of [Li and Lee \(2005\)](#).

We approach the prior distributions of the latent, dynamic effects $\mathbf{K} := (K_{t_{min}}, \dots, K_{t_{max}})$ and $\boldsymbol{\kappa}^{(i)} := (\kappa_{t_{min}}^{(i)}, \dots, \kappa_{t_{max}}^{(i)})$ (with $i = 1, \dots, p$) within the theoretical framework of Gaussian Markov Random Fields ([GMRFs])⁵. GMRFs offer a convenient setting for dealing with conditional (in)dependencies, e.g. in the context of time series analysis (see [Rue and Held \(2005\)](#) and [Congdon \(2006\)](#)). We proceed as follows

$$K_t = \gamma_1 + \gamma_2 t + \rho(K_{t-1} - \gamma_1 - \gamma_2(t-1)) + \sigma_K Z_t, \quad (20)$$

$$\kappa_t^{(i)} = \rho_{(i)} \kappa_{t-1}^{(i)} + \sigma_{\kappa^{(i)}} Z_t^{(i)}, \quad (21)$$

for $i = 1, 2, \dots, p$ and $t = t_{min} + 1, \dots, t_{max}$, where⁶

$$Z_t \sim \mathcal{N}(0, 1) \text{ and } Z_t^{(i)} \sim \mathcal{N}(0, 1).$$

Equivalently, we write

$$K_t | K_{t_{min}}, K_{t_{min}+1}, \dots, K_{t-1} \sim \mathcal{N}(\gamma_1 + \gamma_2 t + \rho(K_{t-1} - \gamma_1 - \gamma_2(t-1)), \sigma_K^2), \quad (22)$$

$$\kappa_t^{(i)} | \kappa_{t_{min}}^{(i)}, \kappa_{t_{min}+1}^{(i)}, \dots, \kappa_{t-1}^{(i)} \sim \mathcal{N}(\rho_{(i)} \kappa_{t-1}^{(i)}, \sigma_{\kappa^{(i)}}^2), \quad (23)$$

for $i = 1, 2, \dots, p$ and $t = t_{min} + 1, \dots, t_{max}$. In addition, we assume the following marginal distributions for $K_{t_{min}}$ and $\kappa_{t_{min}}^{(i)}$ for $i = 1, \dots, p$, following [Rue and Held \(2005\)](#)

$$K_{t_{min}} \sim \mathcal{N}\left(\gamma_1 + \gamma_2 \cdot t_{min}, \frac{\sigma_K^2}{1 - \rho^2}\right) \text{ and } \kappa_{t_{min}}^{(i)} \sim \mathcal{N}\left(0, \frac{\sigma_{\kappa^{(i)}}^2}{1 - \rho_{(i)}^2}\right). \quad (24)$$

Combining (20), (21) and (24) the joint densities of $\mathbf{K}$ and $\boldsymbol{\kappa}^{(i)}$ are

$$\begin{aligned} f(\mathbf{K}) &= f(K_{t_{min}}) \cdot f(K_{t_{min}+1} | K_{t_{min}}) \cdot \dots \cdot f(K_{t_{max}} | K_{t_{max}-1}) \\ &= \frac{1}{(2\pi\sigma_K^2)^{\frac{T}{2}}} |\mathbf{Q}|^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_K^2} (\mathbf{K} - \gamma\mathbf{V}) \mathbf{Q} (\mathbf{K} - \gamma\mathbf{V})'\right), \end{aligned} \quad (25)$$

$$\begin{aligned} f(\boldsymbol{\kappa}^{(i)}) &= f(\kappa_{t_{min}}^{(i)}) \cdot f(\kappa_{t_{min}+1}^{(i)} | \kappa_{t_{min}}^{(i)}) \cdot \dots \cdot f(\kappa_{t_{max}}^{(i)} | \kappa_{t_{max}-1}^{(i)}) \\ &= \frac{1}{(2\pi\sigma_{\kappa^{(i)}}^2)^{\frac{T}{2}}} |\mathbf{Q}^{(i)}|^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_{\kappa^{(i)}}^2} \boldsymbol{\kappa}^{(i)} \mathbf{Q}^{(i)} \boldsymbol{\kappa}^{(i)'}\right), \end{aligned} \quad (26)$$

⁵A random vector $\mathbf{x} := (x_1, \dots, x_n) \in \mathbb{R}^n$ is called a GMRF with mean $\boldsymbol{\mu}$ and symmetric positive-definite precision matrix $\tilde{\mathbf{Q}}$, iff its density has the form,

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\tilde{\mathbf{Q}}|^{\frac{1}{2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}) \tilde{\mathbf{Q}} (\mathbf{x} - \boldsymbol{\mu})'\right).$$

⁶ $\mathcal{N}(\mu, \sigma^2)$ denotes the univariate normal distribution with mean equal to μ and variance equal to σ^2 .

where $\gamma := (\gamma_1 \gamma_2)$, $\mathbf{V} := \begin{pmatrix} \mathbf{1}_T \\ \mathbf{X}_T \end{pmatrix}$, with $\mathbf{1}_T := (1, 1, \dots, 1)$, $\mathbf{X}_T := (1, 2, \dots, T)$, and

$$\mathbf{Q} = \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1 + \rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 + \rho^2 & -\rho \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}, \mathbf{Q}^{(i)} = \begin{pmatrix} 1 & -\rho_{(i)} & 0 & \cdots & 0 \\ -\rho_{(i)} & 1 + \rho_{(i)}^2 & -\rho_{(i)} & \cdots & 0 \\ 0 & -\rho_{(i)} & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 + \rho_{(i)}^2 & -\rho_{(i)} \\ 0 & 0 & \cdots & -\rho_{(i)} & 1 \end{pmatrix}. \quad (27)$$

The representations in (25) and (26) are examples of GMRFs. The matrix $\mathbf{Q}$ is very similar to the precision matrix obtained in Czado et al. (2005). However, due to our assumption (24), we use a different element (1,1) in the precision matrix, namely $\mathbf{Q}_{(1,1)} = 1$ versus $1 + \rho^2$ in Czado et al. (2005). We prefer the specification implied by (25) and (27) because the precision matrices of the common and the population specific effects then have exactly the same structure.

We calculate the corresponding full conditionals, denoted by⁷ $f(K_t | \mathbf{K}_{-t})$ and $f(\kappa_t^{(i)} | \boldsymbol{\kappa}_{-t}^{(i)})$.

- For $t = t_{min}$: $f(K_{t_{min}} | \mathbf{K}_{-t_{min}}) \propto f(\mathbf{K}) \propto f(K_{t_{min}}) \cdot f(K_{t_{min}+1} | K_{t_{min}})$,
 $(K_{t_{min}} | \mathbf{K}_{-t_{min}}) \sim \mathcal{N}(\gamma_1 + \gamma_2 \cdot t_{min} + \rho(K_2 - \gamma_1 - \gamma_2 \cdot (t_{min} + 1)), \sigma_K^2)$;
- For $t \in (t_{min}, t_{max})$: $f(K_t | \mathbf{K}_{-t}) \propto f(\mathbf{K}) \propto f(K_{t+1} | K_t) \cdot f(K_t | K_{t-1})$,
 $(K_t | \mathbf{K}_{-t}) \sim \mathcal{N}\left(\gamma_1 + \gamma_2 \cdot t + \frac{\rho}{1 + \rho^2} (K_{t-1} + K_{t+1} - 2(\gamma_1 + \gamma_2 \cdot t)), \frac{\sigma_K^2}{1 + \rho^2}\right)$;
- For $t = t_{max}$: $f(K_{t_{max}} | \mathbf{K}_{-t_{max}}) \propto f(K_{t_{max}} | K_{t_{max}-1})$,
 $(K_{t_{max}} | \mathbf{K}_{-t_{max}}) \sim \mathcal{N}(\gamma_1 + \gamma_2 \cdot t_{max} + \rho(K_{t_{max}-1} - \gamma_1 - \gamma_2 \cdot (t_{max} - 1)), \sigma_K^2)$.

The corresponding expressions for $f(\kappa_t^{(i)} | \boldsymbol{\kappa}_{-t}^{(i)})$ then follow immediately. We specify a prior distribution for the parameters used in the time series specifications, namely^{8,9}

$$\gamma \sim \mathcal{N}_2(\gamma_0, \Sigma_0), \quad (28)$$

$$\text{logit}(\rho) \sim \mathcal{N}(\mu_\rho, \sigma_\rho^2), \quad (29)$$

$$\text{logit}(\rho_{(i)}) \sim \mathcal{N}(\mu_{\rho_{(i)}}, \sigma_{\rho_{(i)}}^2), \quad (30)$$

where γ_0 is 1×2 vector and Σ_0 is a 2×2 matrix of constants. The logit transformation ensures that the hyperparameters ρ and $\rho_{(i)}$ are positive and less than one, see Cairns et al. (2011). We follow the standard practice, see Gelman et al. (2013), and choose an Inverse Gamma ($[\mathcal{IG}]$) prior for σ_K^2 and $\sigma_{\kappa^{(i)}}^2$. This implies¹⁰

$$\sigma_K^{-2} \sim \mathcal{G}(a_K, b_K) \text{ and } \sigma_{\kappa^{(i)}}^{-2} \sim \mathcal{G}(a_{\kappa^{(i)}}, b_{\kappa^{(i)}}).$$

⁷ $\mathbf{K}_{-t} := (K_{t_{min}}, K_2, \dots, K_{t-1}, K_{t+1}, \dots, K_{t_{max}})$ and $\boldsymbol{\kappa}_{-t}^{(i)} := (\kappa_{t_{min}}^{(i)}, \kappa_2^{(i)}, \dots, \kappa_{t-1}^{(i)}, \kappa_{t+1}^{(i)}, \dots, \kappa_{t_{max}}^{(i)})$. To save space we do not mention the parameters used in the distribution of K_t , see Section 4 for full details.

⁸ $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the m -dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

⁹ The density of a logit normal distribution is $f(x; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \frac{1}{x(1-x)} \exp\left(-\frac{(\text{logit}(x) - \mu)^2}{2\sigma^2}\right)$, $x \in (0, 1)$.

¹⁰ $\mathcal{G}$ denotes the Gamma distribution, with density $f(x; a, b) := \frac{1}{\Gamma(a)} b^a x^{a-1} e^{-xb}$, $x \in (0, +\infty)$, mean equal to $a \cdot b^{-1}$ and variance equal to $a \cdot b^{-2}$.

3.2 Prior distributions for the age parameters

Following [Czado et al. \(2005\)](#) we use the following prior for the $\beta_{x,j}^{(i)}$'s (with $j = 1, 2$) in the LC-2,t model

$$\beta_j^{(i)} \sim \mathcal{N}_M \left(\frac{1}{M} \mathbf{1}_M, \sigma_{\beta_j^{(i)}}^2 \mathbf{I}_M \right). \quad (31)$$

Hereby $M = x_{\max} - x_{\min} + 1$ is the number of ages in the sample, $\beta_j^{(i)} := (\beta_{x_{\min},j}^{(i)}, \dots, \beta_{x_{\max},j}^{(i)})$ for $j = 1, 2$, $\mathbf{1}_M$ is the $1 \times M$ unit vector and $\mathbf{I}_M$ the $M \times M$ identity matrix. For the variance hyperparameters we have

$$\sigma_{\beta_j^{(i)}}^{-2} \sim \mathcal{G} \left(a_{\beta_j^{(i)}}^{(i)}, b_{\beta_j^{(i)}}^{(i)} \right). \quad (32)$$

For the LL model we use (31) as the prior for $\beta_x^{(i)}$ and the common age-dependent parameter B_x has prior

$$\mathbf{B} \sim \mathcal{N}_M \left(\frac{1}{M} \mathbf{1}_M, \sigma_B^2 \mathbf{I}_M \right), \quad (33)$$

with $\mathbf{B} := (B_{x_{\min}}, \dots, B_{x_{\max}})$. The prior distribution for the variance parameter is,

$$\sigma_B^{-2} \sim \mathcal{G} (a_B, b_B). \quad (34)$$

We adopt the prior distribution of [Czado et al. \(2005\)](#) for $\exp(\alpha_x^{(i)})$, namely

$$e_x^{(i)} := \exp(\alpha_x^{(i)}) \sim \mathcal{G} (a_x^{(i)}, b_x^{(i)}), \quad (35)$$

where $a_x^{(i)}$ and $b_x^{(i)}$ are prespecified constants, with $i = 1, 2, \dots, p$ and $x = x_{\min}, \dots, x_{\max}$. For the LL model we use (35) as the prior for $\alpha_x^{(i)}$ and for the common parameter A_x we have

$$\mathcal{E}_x := \exp(A_x) \sim \mathcal{G} (a_x, b_x). \quad (36)$$

4 Posterior distributions

We include a detailed outline of the posterior distributions used in the LC-2,t model and refer to an online appendix for the posterior distributions of the LL model. Inference about the parameters used in the model, given the observed deaths and exposures, follows from Bayes' theorem¹¹

$$f(\mathbf{K}, \boldsymbol{\kappa}, \beta_1, \beta_2, \boldsymbol{\alpha} | \mathbf{D}, \mathbf{E}) \propto f(\mathbf{D}, \mathbf{E} | \mathbf{K}, \boldsymbol{\kappa}, \beta_1, \beta_2, \boldsymbol{\alpha}) \cdot f(\mathbf{K}, \boldsymbol{\kappa}, \beta_1, \beta_2, \boldsymbol{\alpha}).$$

We specify the technical details of the MCMC algorithm. In particular, we combine Gibbs sampling and, if necessary, Metropolis-Hastings steps. We introduce the following notation for the set of data, latent effects and (hyper)parameters

$$\boldsymbol{\theta} := \{ \mathbf{D}, \mathbf{E}, \mathbf{K}, \boldsymbol{\kappa}, \beta_1, \beta_2, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\rho}, \sigma_K^2, \sigma_{\kappa}^2, \sigma_{\beta_1}^2, \sigma_{\beta_2}^2 \},$$

where $\boldsymbol{\rho} := (\rho, \rho_{(1)}, \dots, \rho_{(p)})$, $\sigma_{\kappa}^2 := (\sigma_{\kappa(1)}^2, \dots, \sigma_{\kappa(p)}^2)$ and $\sigma_{\beta_j}^2 := (\sigma_{\beta_j^{(1)}}^2, \dots, \sigma_{\beta_j^{(p)}}^2)$.

¹¹Notation used and not introduced before: $\alpha^{(i)} := (\alpha_{x_{\min}}^{(i)}, \dots, \alpha_{x_{\max}}^{(i)})$, $\boldsymbol{\alpha} := (\boldsymbol{\alpha}^{(1)}, \dots, \boldsymbol{\alpha}^{(p)})$, $\beta_j := (\beta_j^{(1)}, \dots, \beta_j^{(p)})$ and $\boldsymbol{\kappa} := (\boldsymbol{\kappa}^{(1)}, \dots, \boldsymbol{\kappa}^{(p)})$, $\mathbf{D}^{(i)}$ respectively $\mathbf{E}^{(i)}$ are the $M \times T$ matrices of deaths and exposures respectively, for the i -th population. Hereby, $M = x_{\max} - x_{\min} + 1$ the number of ages in the sample, and $T = t_{\max} - t_{\min} + 1$ is the number of years in the calibration period. We denote the partitioned matrices $\mathbf{D} := (\mathbf{D}^{(1)}, \dots, \mathbf{D}^{(p)})$ and $\mathbf{E} := (\mathbf{E}^{(1)}, \dots, \mathbf{E}^{(p)})$.

4.1 Metropolis-Hastings sampling for κ 's and K

For the period effect K_t the proportional full conditional density function is

$$\begin{aligned} f(K_t | \boldsymbol{\theta} \setminus \{K_t\}) &= f(K_t | \mathbf{D}, \mathbf{E}, \mathbf{K}_{-t}, \boldsymbol{\kappa}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \rho, \sigma_K^2) \\ &\propto f(\mathbf{D}_{:,t}, \mathbf{E}_{:,t} | K_t, \boldsymbol{\kappa}_t, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}) \cdot f(K_t | \mathbf{K}_{-t}, \boldsymbol{\gamma}, \rho, \sigma_K^2), \end{aligned} \quad (37)$$

where $\boldsymbol{\kappa}_t := (\kappa_t^{(1)}, \dots, \kappa_t^{(p)})$. For the random vectors $\mathbf{D}_{:,t} := (\mathbf{D}_{:,t}^{(1)}, \dots, \mathbf{D}_{:,t}^{(p)})$, with $\mathbf{D}_{:,t}^{(i)} := (D_{x_{\min},t}^{(i)}, \dots, D_{x_{\max},t}^{(i)})$ ¹², we have

$$\begin{aligned} f(\mathbf{D}_{:,t}, \mathbf{E}_{:,t} | K_t, \boldsymbol{\kappa}_t, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}) &= \prod_{i=1}^p f(\mathbf{D}_{:,t}^{(i)}, \mathbf{E}_{:,t}^{(i)} | K_t, \kappa_t^{(i)}, \beta_1^{(i)}, \beta_2^{(i)}, \boldsymbol{\alpha}^{(i)}) \\ &\propto \prod_{i=1}^p \prod_{x=x_{\min}}^{x_{\max}} \exp\left(-E_{x,t}^{(i)} \exp\left(\alpha_x^{(i)} + \beta_{x,1}^{(i)} K_t + \beta_{x,2}^{(i)} \kappa_t^{(i)}\right)\right) \\ &\quad \times \exp\left(\left[\beta_{x,1}^{(i)} K_t + \beta_{x,2}^{(i)} \kappa_t^{(i)}\right] D_{x,t}^{(i)}\right). \end{aligned} \quad (38)$$

In a similar way we obtain for $\kappa_t^{(i)}$

$$f(\kappa_t^{(i)} | \boldsymbol{\theta} \setminus \{\kappa_t^{(i)}\}) \propto f(\mathbf{D}_{:,t}^{(i)}, \mathbf{E}_{:,t}^{(i)} | K_t, \kappa_t^{(i)}, \beta_1^{(i)}, \beta_2^{(i)}, \boldsymbol{\alpha}^{(i)}) \cdot f(\kappa_t^{(i)} | \boldsymbol{\kappa}_{-t}^{(i)}, \rho^{(i)}, \sigma_{\kappa^{(i)}}^2). \quad (39)$$

The MH steps for updating the K_t 's are listed in Appendix A.1. Updating the $\kappa_t^{(i)}$'s for each population i goes in a similar way.

4.2 Metropolis-Hastings sampling for β 's

The proportional conditional probability density function of the first age effect $\beta_{x,1}^{(i)}$ is

$$f(\beta_{x,1}^{(i)} | \boldsymbol{\theta} \setminus \{\beta_{x,1}^{(i)}\}) \propto f(\mathbf{D}_{x,\cdot}^{(i)}, \mathbf{E}_{x,\cdot}^{(i)} | \mathbf{K}, \boldsymbol{\kappa}^{(i)}, \beta_{x,1}^{(i)}, \beta_{x,2}^{(i)}, \alpha_x^{(i)}) \cdot f(\beta_{x,1}^{(i)} | \sigma_{\beta_1^{(i)}}^2). \quad (40)$$

For the vector¹³ $\mathbf{D}_{x,\cdot}^{(i)} := (D_{x,t_{\min}}^{(i)}, \dots, D_{x,t_{\max}}^{(i)})$ we have

$$\begin{aligned} f(\mathbf{D}_{x,\cdot}^{(i)}, \mathbf{E}_{x,\cdot}^{(i)} | \mathbf{K}, \boldsymbol{\kappa}^{(i)}, \beta_{x,1}^{(i)}, \beta_{x,2}^{(i)}, \alpha_x^{(i)}) &\propto \prod_{t=t_{\min}}^{t_{\max}} \exp\left(-E_{x,t}^{(i)} \exp\left(\alpha_x^{(i)} + \beta_{x,1}^{(i)} K_t + \beta_{x,2}^{(i)} \kappa_t^{(i)}\right)\right) \\ &\quad \times \exp\left(\left[\beta_{x,1}^{(i)} K_t + \beta_{x,2}^{(i)} \kappa_t^{(i)}\right] D_{x,t}^{(i)}\right). \end{aligned}$$

The corresponding MH steps are given in Appendix A.2. The updating scheme for $\beta_{x,2}^{(i)}$ follows immediately.

¹²We use the notation $\mathbf{E}_{:,t} := (\mathbf{E}_{:,t}^{(1)}, \dots, \mathbf{E}_{:,t}^{(p)})$ and $\mathbf{E}_{:,t}^{(i)} := (E_{x_{\min},t}^{(i)}, \dots, E_{x_{\max},t}^{(i)})$

¹³Similarly we denote $\mathbf{E}_{x,\cdot}^{(i)} := (E_{x,t_{\min}}^{(i)}, \dots, E_{x,t_{\max}}^{(i)})$

4.3 Gibbs sampling for α 's

We simulate directly from the posterior distribution of the $\alpha^{(i)}$ parameters because the Gamma distribution in (35) is a conjugate prior for the Poisson likelihood function. With $\mathbf{e}^{(i)} := (e_{x_{\min}}^{(i)}, e_{x_{\min}+1}^{(i)}, \dots, e_{x_{\max}}^{(i)})$ and $e_x^{(i)} = \exp(\alpha_x^{(i)})$ we have

$$f(\mathbf{D}^{(i)}, \mathbf{E}^{(i)} | \mathbf{K}, \boldsymbol{\kappa}^{(i)}, \beta_1^{(i)}, \beta_2^{(i)}, e_x^{(i)}) \propto \exp(-c_x^{(i)} e_x^{(i)}) (e_x^{(i)})^{D_{x,\bullet}^{(i)}}, \quad (41)$$

where $c_x^{(i)} := \sum_{t=t_{\min}}^{t_{\max}} E_{x,t}^{(i)} \exp(\beta_{x,1}^{(i)} K_t + \beta_{x,2}^{(i)} \kappa_t^{(i)})$ and $D_{x,\bullet}^{(i)} := \sum_{t=t_{\min}}^{t_{\max}} D_{x,t}^{(i)}$. As a result the posterior distribution of $e_x^{(i)}$ is again Gamma distributed, as follows

$$\begin{aligned} f(e_x^{(i)} | \boldsymbol{\theta} \setminus \{e_x^{(i)}\}) &\propto f(\mathbf{D}^{(i)}, \mathbf{E}^{(i)} | \mathbf{K}, \boldsymbol{\kappa}^{(i)}, \beta_1^{(i)}, \beta_2^{(i)}, e_x^{(i)}) \cdot f(e_x^{(i)}) \\ &\propto \exp\left(-\left(b_x^{(i)} + c_x^{(i)}\right) e_x^{(i)}\right) (e_x^{(i)})^{a_x^{(i)} + D_{x,\bullet}^{(i)} - 1} \\ &\sim \mathcal{G}\left(a_x^{(i)} + D_{x,\bullet}^{(i)}, b_x^{(i)} + c_x^{(i)}\right). \end{aligned} \quad (42)$$

4.4 Metropolis-Hastings sampling for ρ and $\rho_{(i)}$

Czado et al. (2005) derive an explicit expression for the posterior distribution of ρ in their trend stationary LC model. However, in our setting the posterior distributions for ρ and $\rho^{(i)}$ do not follow a standard distribution and we update these parameters using MH steps. Denote $\tilde{K}_t := K_t - \gamma_1 - \gamma_2 t$ and $\tilde{\mathbf{Q}} := \sigma_K^{-2} \mathbf{Q}$, with determinant $|\tilde{\mathbf{Q}}| = \sigma_K^{-2T} (1 - \rho^2)$, then

$$\begin{aligned} f(\rho | \boldsymbol{\theta} \setminus \{\rho\}) &= f(\rho | \mathbf{K}, \gamma, \sigma_K^2) \propto f(\mathbf{K} | \gamma, \rho, \sigma_K^2) \cdot f(\rho) \\ &\propto |\tilde{\mathbf{Q}}|^{\frac{1}{2}} \exp\left(-\frac{1}{2} (\mathbf{K} - \gamma \mathbf{V})' \tilde{\mathbf{Q}} (\mathbf{K} - \gamma \mathbf{V})\right) \frac{1}{\rho(1-\rho)} \exp\left(-\frac{(\text{logit}(\rho) - \mu_\rho)^2}{2\sigma_\rho^2}\right) \\ &\propto \exp\left(-\frac{1}{2\sigma_K^2} \left[\rho^2 \sum_{t=t_{\min}+1}^{t_{\max}-1} \tilde{K}_t^2 - 2\rho \sum_{t=t_{\min}+1}^{t_{\max}} \tilde{K}_t \tilde{K}_{t-1} \right]\right) \frac{(1-\rho^2)^{\frac{1}{2}}}{\rho(1-\rho)} \exp\left(-\frac{(\text{logit}(\rho) - \mu_\rho)^2}{2\sigma_\rho^2}\right) \\ &\propto \exp\left(-\frac{1}{2\sigma_K^2 a_\rho^{-1}} \left(\rho - \frac{b_\rho}{a_\rho}\right)^2\right) \frac{1}{\rho} \left(\frac{1+\rho}{1-\rho}\right)^{\frac{1}{2}} \exp\left(-\frac{(\text{logit}(\rho) - \mu_\rho)^2}{2\sigma_\rho^2}\right), \end{aligned}$$

where $a_\rho := \sum_{t=t_{\min}+1}^{t_{\max}-1} \tilde{K}_t^2$ and $b_\rho := \sum_{t=t_{\min}+1}^{t_{\max}} \tilde{K}_t \tilde{K}_{t-1}$. Full details of the MH steps to update this parameter are presented in Appendix A.3. The MH steps for the population specific $\rho_{(i)}$ follow immediately, using

$$\begin{aligned} f(\rho_{(i)} | \boldsymbol{\theta} \setminus \{\rho_{(i)}\}) \\ \propto \exp\left(-\frac{1}{2\sigma_{\kappa(i)}^2 a_{\rho(i)}^{-1}} \left(\rho_{(i)} - \frac{b_{\rho(i)}}{a_{\rho(i)}}\right)^2\right) \frac{1}{\rho_{(i)}} \left(\frac{1+\rho_{(i)}}{1-\rho_{(i)}}\right)^{\frac{1}{2}} \exp\left(-\frac{(\text{logit}(\rho_{(i)}) - \mu_{\rho(i)})^2}{2\sigma_{\rho(i)}^2}\right), \end{aligned}$$

where $a_{\rho(i)} = \sum_{t=t_{\min}+1}^{t_{\max}-1} (\kappa_t^{(i)})^2$ and $b_{\rho(i)} = \sum_{t=t_{\min}+1}^{t_{\max}} \kappa_t^{(i)} \kappa_{t-1}^{(i)}$.

4.5 Gibbs sampling for γ_1 and γ_2

The posterior distribution of γ becomes

$$\begin{aligned} f(\gamma|\boldsymbol{\theta} \setminus \{\gamma\}) &\propto f(\mathbf{K}|\gamma, \rho, \sigma_K^2) \cdot f(\gamma) \\ &\propto \exp\left(-\frac{1}{2}(\mathbf{K} - \gamma\mathbf{V})\tilde{\mathbf{Q}}(\mathbf{K} - \gamma\mathbf{V})'\right) \exp\left(-\frac{1}{2}(\gamma - \gamma_0)\Sigma_0^{-1}(\gamma - \gamma_0)'\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\gamma\left[\mathbf{V}\tilde{\mathbf{Q}}\mathbf{V}' + \Sigma_0^{-1}\right]\gamma' - 2\gamma\left[\mathbf{V}\tilde{\mathbf{Q}}\mathbf{K} + \Sigma_0^{-1}\gamma_0'\right]\right)\right). \end{aligned}$$

Thus,

$$(\gamma|\boldsymbol{\theta} \setminus \{\gamma\}) \sim \mathcal{N}_2(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*),$$

where $\boldsymbol{\Sigma}_* := (\mathbf{V}\tilde{\mathbf{Q}}\mathbf{V}' + \Sigma_0^{-1})^{-1}$ and $\boldsymbol{\mu}_* := \boldsymbol{\Sigma}_* (\mathbf{V}\tilde{\mathbf{Q}}\mathbf{K} + \Sigma_0^{-1}\gamma_0')$.

4.6 Gibbs sampling for σ 's

The posterior distribution of the variance parameter σ_K^2 is

$$\begin{aligned} f(\sigma_K^2|\boldsymbol{\theta} \setminus \{\sigma_K^2\}) &\propto f(\mathbf{K}|\gamma, \rho, \sigma_K^2) \cdot f(\sigma_K^2) \\ &\propto \sigma_K^{-T} \exp\left(-\frac{1}{2}(\mathbf{K} - \gamma\mathbf{V})\mathbf{Q}(\mathbf{K} - \gamma\mathbf{V})'\right) \sigma_K^{-2(a_K-1)} \exp(-\sigma_K^{-2}b_K) \\ &\propto \sigma_K^{-2(a_K+\frac{T}{2}-1)} \exp\left(-\sigma_K^{-2}\left(b_K + \frac{1}{2}(\mathbf{K} - \gamma\mathbf{V})\mathbf{Q}(\mathbf{K} - \gamma\mathbf{V})'\right)\right). \end{aligned}$$

Consequently, the posterior distribution is

$$(\sigma_K^2|\boldsymbol{\theta} \setminus \{\sigma_K^2\}) \sim \mathcal{IG}\left(a_K + \frac{T}{2}, b_K + \frac{1}{2}(\mathbf{K} - \gamma\mathbf{V})\mathbf{Q}(\mathbf{K} - \gamma\mathbf{V})'\right). \quad (43)$$

Analogously, we obtain

$$\begin{aligned} (\sigma_{\kappa^{(i)}}^2|\boldsymbol{\theta} \setminus \{\sigma_{\kappa^{(i)}}^2\}) &\sim \mathcal{IG}\left(a_{\kappa^{(i)}} + \frac{T}{2}, b_{\kappa^{(i)}} + \frac{1}{2}\boldsymbol{\kappa}^{(i)}\mathbf{Q}\boldsymbol{\kappa}^{(i)'}\right), \\ (\sigma_{\beta_j^{(i)}}^2|\boldsymbol{\theta} \setminus \{\sigma_{\beta_j^{(i)}}^2\}) &\sim \mathcal{IG}\left(a_{\beta_j^{(i)}} + \frac{M}{2}, b_{\beta_j^{(i)}} + \frac{1}{2}\boldsymbol{\beta}_j^{(i)}\boldsymbol{\beta}_j^{(i)'}\right). \end{aligned}$$

5 Empirical studies

We illustrate the multi-population LC-2,t and LL models in two applications. In the first application we model Swedish mortality data for males and females (as in [Li and Lee \(2005\)](#) and [Hyndman et al. \(2013\)](#)) and the second illustration analyzes gender specific mortality data for a collection of 14 European countries (as in [Koninklijk Actuarieel Genootschap \(2014\)](#)).

5.1 Swedish male and female mortality

Data. We use Swedish mortality data for males and females, available from the Human Mortality Database ([HMD])^{14,15}. We calibrate the model on data from 1950-2009, for ages 0-89. [Li and Lee \(2005\)](#) and [Hyndman et al. \(2013\)](#) analyze the same data set with the augmented common factor model and the product-ratio functional time series analysis approach, respectively. Figures 1 and 2 in [Hyndman et al. \(2013\)](#) (pages 268-269) visualize the evolution of Swedish mortality data over age and time using rainbow plots¹⁶.

Initializing the prior distributions. To initialize the prior distributions proposed in Section 3 we have to fix the constants used in these specifications. We show results calculated with the following set of constants. Hereby i refers to females ('F') or males ('M').

- In the LC-2,t model we use $a_x^{(i)} = b_x^{(i)} \cdot \exp\{\hat{\alpha}_x^{(i)}\}$ where $b_x^{(i)} = 0.01$ and

$$\hat{\alpha}_x^{(i)} = \frac{1}{T} \sum_{t=t_{\min}}^{t_{\max}} \log \left(\frac{d_{x,t}^{(i)}}{E_{x,t}^{(i)}} \right). \quad (44)$$

Hence $\mathbb{E}(\exp\{\alpha_x^{(i)}\}) = \exp\{\hat{\alpha}_x^{(i)}\}$. For the LL model we use $a_x = b_x \cdot \exp\{\hat{A}_x\}$, where $b_x = 0.01$ and

$$\hat{A}_x = \frac{1}{T} \sum_{t=t_{\min}}^{t_{\max}} \log \left(\frac{d_{x,t}^{(\bullet)}}{E_{x,t}^{(\bullet)}} \right), \quad (45)$$

and for the population specific $\alpha_x^{(i)}$ parameters we use $a_x^{(i)} = b_x^{(i)} \cdot \exp\{\hat{\alpha}_x^{(i)}\}$, where $b_x^{(i)} = 1$ and

$$\hat{\alpha}_x^{(i)} = \frac{1}{T} \sum_{t=t_{\min}}^{t_{\max}} \log \left(\frac{d_{x,t}^{(i)}}{E_{x,t}^{(i)}} \right) - \hat{A}_x. \quad (46)$$

Consequently, $\mathbb{E}(\exp\{\alpha_x^{(i)}\}) = \exp\{\hat{\alpha}_x^{(i)}\}$. As such, we apply the strategy of [Czado et al. \(2005\)](#) to initialize the priors of the age specific intercepts.

- To obtain meaningful values for γ_0 we fit the LC model, using Maximum Likelihood Estimation with the Poisson likelihood assumption, to $D_{x,t}^{(\bullet)}$ and $E_{x,t}^{(\bullet)}$. The resulting period effects are denoted by $\hat{K}_t^{\text{LC}}$. We fit a least squares line through coordinates $(t, \hat{K}_t^{\text{LC}})$ (with $t = t_{\min}, \dots, t_{\max}$) and use its intercept and slope as γ_0 and γ_1 respectively. The Σ_0 matrix is a diagonal matrix with diagonal elements 1 and 1 respectively, which results in a vague prior. We use these specifications for both multi-population mortality models in the study.
- We follow [Cairns et al. \(2011\)](#) and set $\mu_\rho = 3$ and $\sigma_\rho^2 = 0.5^2$. For the corresponding country specific hyperparameters we use $\mu_{\rho^{(i)}} = 0.5$ and $\sigma_{\rho^{(i)}}^2 = 0.5^2$ for $i = 1, 2, \dots, p$.

¹⁴See www.mortality.org.

¹⁵More specifically, we use 1×1 (Age interval $\times$ Year interval) tables of Deaths and Exposures-to-risk.

¹⁶See R package [grDevices](#).

- Inspired by [Czado et al. \(2005\)](#), we choose¹⁷ $a_K = a_{\kappa}^{(i)} = 2.1$ and $b_K = b_{\kappa}^{(i)} = 1$, thus $\mathbb{E}(\sigma_K^2) = \mathbb{E}(\sigma_{\kappa(i)}^2) \approx 0.9$ and $\text{Var}(\sigma_K^2) = \text{Var}(\sigma_{\kappa(i)}^2) \approx 8.3$. For the $\beta_{x,1}^{(i)}$ parameters we choose $a_{\beta_1}^{(i)} = 2.1$ and $b_{\beta_1}^{(i)} = (a_{\beta_1}^{(i)} - 1) \cdot \text{Var}(\hat{B}_x^{\text{LC}})$ for all i , where $\hat{B}_x^{\text{LC}}$ is obtained from the LC model estimated on the aggregated data. For $\beta_{x,2}^{(i)}$ we use $a_{\beta_2}^{(i)} = 2.1$ and $b_{\beta_2}^{(i)} = 0.1$, such that $\mathbb{E}(\sigma_{\beta_2}^2) \approx 0.09$ and $\text{Var}(\sigma_{\beta_2}^2) \approx 0.83$.

Convergence diagnostics. We run 20,000 iterations of the MCMC algorithm for the LC-2,t model. We remove an initial, burn-in period of 4,000 iterations, and collect information from the remaining iterations, using a thinning factor of 10. We perform the usual convergence checks. The trace plots show good mixing properties¹⁸. For the LL model we first update the common parameters as described in the online appendix to this paper. After reaching convergence of the MCMC updating scheme for the common parameters, we simulate the population specific parameters from their posterior distributions, while replacing in each iteration the common parameters in the likelihood with a value sampled from their posterior distributions. Our final sample size is again 1,600.

Estimates and credibility intervals in the LC-2,t model. Figure 1 shows the parameter estimates of $\alpha_x^{(M)}$, $\alpha_x^{(F)}$ (top row), $\beta_{x,1}^{(M)}$, $\beta_{x,1}^{(F)}$ (second row) and $\beta_{x,2}^{(M)}$ and $\beta_{x,2}^{(F)}$ (bottom row) as well as their 95% credibility intervals. The Bayesian estimates of the α_x parameters are very similar to the age specific intercepts estimated with a single population LC model. The Bayesian estimates of $\beta_{x,1}^{(M)}$ and $\beta_{x,1}^{(F)}$ have a similar course, though for ages around 25 and 75, $\beta_{x,1}^{(M)}$ and $\beta_{x,1}^{(F)}$ show some clear differences.

The samples from the posterior distribution of the common period effect K_t , displayed in the left panel of Figure 2, show mean reversion around a linear, downward trend with the posterior median of ρ equal to 0.947. The intercept of the linear trend has posterior median equal to 48.88 and the posterior median of the slope is -1.589 . The gender specific latent effects $\kappa_t^{(F)}$ and $\kappa_t^{(M)}$ are in the right panel of Figure 2. These show mean reversion around 0 with the posterior median of $\rho^{(M)}$ equal to 0.90 and the posterior median of $\rho^{(F)}$ is 0.71. We compare the Bayesian estimates of K_t to the trend obtained with MLE in the Poisson LC model on $D_{x,t}^{(\bullet)}$ and $E_{x,t}^{(\bullet)}$ and with the common trend estimated using SVD in the LL model. The Bayesian common trend in the LC-2,t model is very similar to, though visibly different from, these two benchmark estimates.

Estimates and credibility intervals in the LL model. Figure 3 shows the parameter estimates and the 95% credibility intervals of $\alpha_x^{(M)}$, $\alpha_x^{(F)}$, A_x and Figure 4 shows corresponding results for $\beta_x^{(M)}$, $\beta_x^{(F)}$, B_x in the Bayesian LL model. As expected, A_x and B_x capture the common pattern in the age specific intercepts and slopes shown in Figure 1.

Estimates and credibility intervals of the common latent effect, K_t , and the gender specific period effects, $\kappa_t^{(F)}$ and $\kappa_t^{(M)}$, are in Figure 5. The common trend shows a downward, rather linear pattern, mean reverting around a linear trend with the posterior median of ρ equal to

¹⁷If $\sigma^2 \sim \mathcal{IG}(a, b)$ then $\mathbb{E}(\sigma^2) = \frac{b}{a-1}$ and $\text{Var}(\sigma^2) = \frac{b^2}{(a-1)^2(a-2)}$.

¹⁸Trace plots of the (hyper)parameters used in the model are available in a note accompanying this paper, downloadable from www.econ.kuleuven.be/anastasios.bardoutsos.

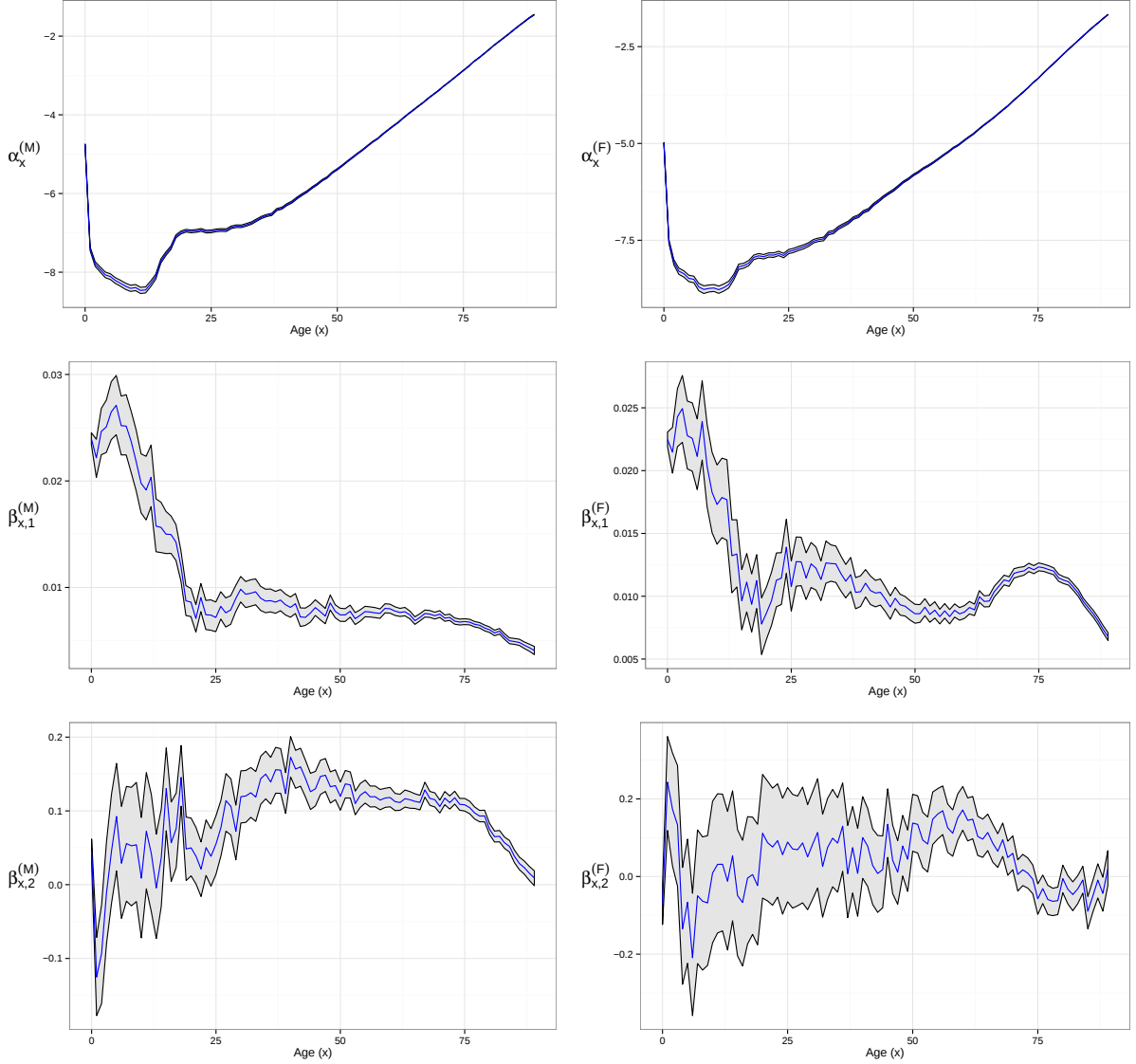


Figure 1: Model LC-2,t: parameters $\alpha_x^{(M)}$, $\alpha_x^{(F)}$ (top row), $\beta_{x,1}^{(M)}$, $\beta_{x,1}^{(F)}$ (second row), $\beta_{x,2}^{(M)}$ and $\beta_{x,2}^{(F)}$ (bottom row), calibrated on Swedish male and female mortality data, period 1950-2009 and ages 0-89: 20,000 iterations, burn-in of 4,000 iterations and thinning factor 10. The blue line represents the median of the 1,600 iterations kept after thinning and burn-in. The shaded gray area is the corresponding 95% pointwise credibility interval.

0.958. The linear trend has an intercept with posterior median equal to 48.88 and the posterior median of the slope is -1.66 . The Bayesian estimate of the common trend is now very close to the common trend estimated by MLE in the Poisson LC model calibrated on $D_{x,t}^{(\bullet)}$ and $E_{x,t}^{(\bullet)}$. The gender specific latent effects $\kappa_t^{(F)}$ and $\kappa_t^{(M)}$ are in the right panel of Figure 5. These show mean reversion around 0, as specified in their time series prior, with the posterior median of $\rho^{(M)}$ equal to 0.908 and the posterior median of $\rho^{(F)}$ is 0.9.

Forecasting mortality rates. Figures 6 and 7 show the Bayesian multi-population estimates and projections of the force of mortality, $\mu_{x,t}^{(M)}$ and $\mu_{x,t}^{(F)}$, and their corresponding credibility

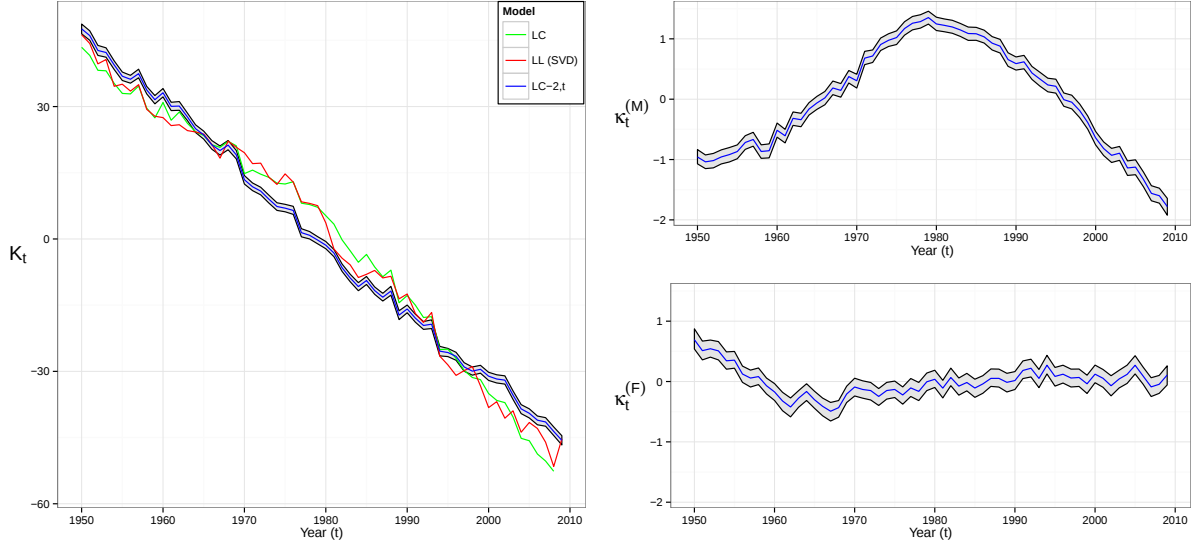


Figure 2: Model LC-2,t: latent effects K_t , $\kappa_t^{(M)}$ and $\kappa_t^{(F)}$, calibrated on Swedish male and female mortality data, period 1950-2009 and ages 0-89: 20,000 iterations, burn-in of 4,000 iterations and thinning factor 10. The blue line represents the median calculated from the 1,600 iterations kept after thinning and burn-in. The shaded gray area is the corresponding 95% pointwise credibility interval. The green reference line is $\hat{K}_t^{LC}$ and the red reference line is the common latent factor, obtained with SVD in the LL model.

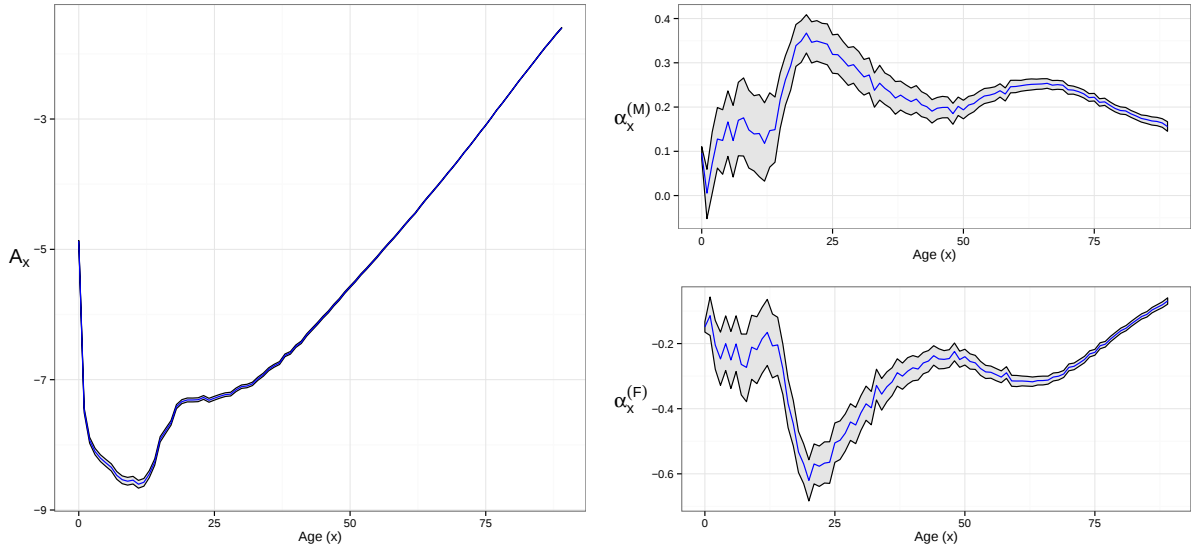


Figure 3: Model LL: parameters $\alpha_x^{(M)}$, $\alpha_x^{(F)}$ and A_x , calibrated on Swedish mortality data, period 1950-2009 and ages 0-89: results based on 1,600 iterations after burn-in and thinning. The blue line represents the median of these 1,600 iterations. The shaded gray area is the 95% pointwise credibility interval using these iterations.

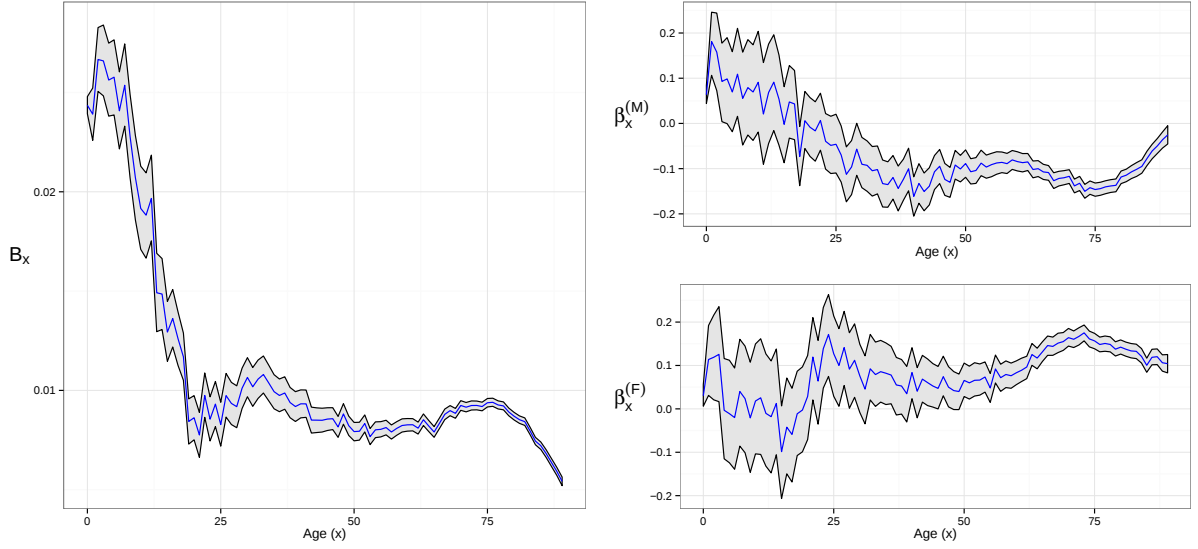


Figure 4: Model LL: parameters $\beta_x^{(M)}$, $\beta_x^{(F)}$ and B_x , calibrated on Swedish mortality data, period 1950-2009 and ages 0-89: results based on 1,600 iterations after burn-in and thinning. The blue line represents the median of the 1,600 iterations. The shaded gray area is the 95% pointwise credibility interval using these iterations.

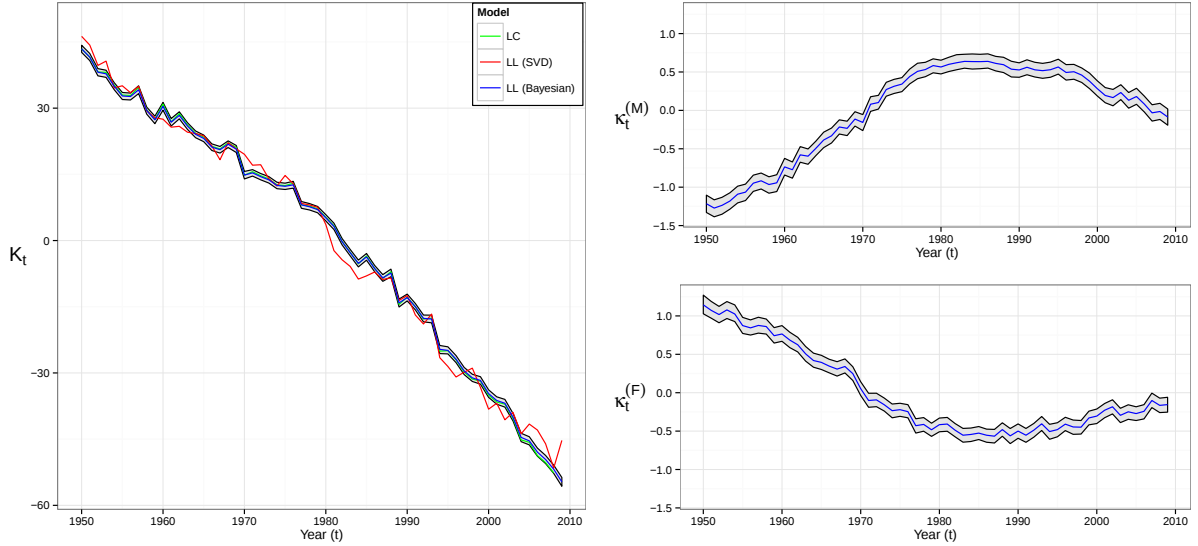


Figure 5: Model LL: latent effects K_t , $\kappa_t^{(M)}$ and $\kappa_t^{(F)}$, calibrated on Swedish mortality data, period 1950-2009 and ages 0-89: results based on 1,600 iterations after burn-in and thinning. The blue line represents the median of these 1,600 iterations. The shaded gray area is the 95% pointwise credibility interval using these iterations. The green reference line is $\hat{K}_t^{LC}$ and the red reference line is the common latent factor, obtained with SVD in the LL model.

intervals for a selection of ages, namely the 20, 40, 60 and 80 years old. We benchmark these results against the estimates obtained with a Poisson MLE implementation of a gender specific LC model and the LL model calibrated using the Singular Value Decomposition. These benchmark models use a random walk with drift for the period effect in the LC model and the common time trend in the LL model, as well as an AR(1) model for the $\kappa_t^{(i)}$ in the LL model. Parameters in these time series models are estimated with MLE. The multi-population models reveal better in sample fitting compared to the gender specific LC models, as illustrated by, for example, graph (b) in Figure 6. Both Bayesian multi-population models follow the observed data closely for all ages in the sample, whereas the SVD calibration of LL for some ages is not capturing well the observed evolution. Graphs (d) in Figure 6 and 7 illustrate this. Graph (b) in Figure 6 and (c) in Figure 7 illustrate that the LC-2,t model is following the observed data slightly more closely than the LL model, due to the extra flexibility implied by using age and gender specific slopes $\beta_{x,1}^{(i)}$.

5.2 A set of European countries

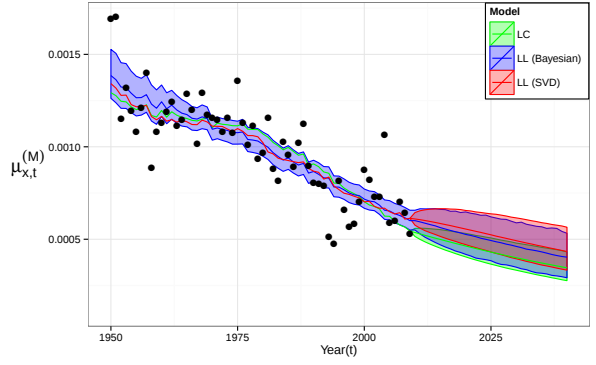
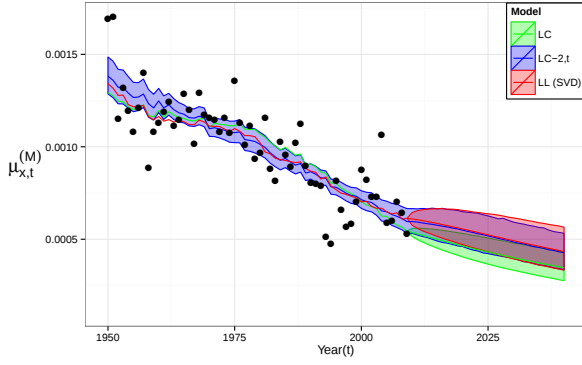
Data. We select fourteen European countries, namely Austria, Belgium, Denmark, England-Wales, France, West Germany, Luxembourg, Netherlands, Switzerland, Finland, Iceland, Ireland, Norway, and Sweden, and analyze mortality data for females as well as males from the HMD database. The Royal Dutch Actuarial Association¹⁹ uses this set of countries in its most recent stochastic projection model, see [Koninklijk Actuariel Genootschap \(2014\)](#). This collection of countries meets the requirements formulated by [Li and Lee \(2005\)](#) for the study of mortality rates as a group, i.e. they have similar socio-economic conditions and close connections which are expected to continue in the future. We use ages from 0 until 89, as in [Pitacco et al. \(2009\)](#). Fitted values and forecasts for older ages can be obtained with a technique to close mortality tables, see for example [Denuit and Goderniaux \(2005\)](#) and [Pitacco et al. \(2009\)](#) (Section 2.8). The calibration period in this illustration is the interval [1975, 2009], which is in line with the optimal calibration period obtained for Belgian data in [Pitacco et al. \(2009\)](#) (Section 5.5.3) using the technique of [Denuit and Goderniaux \(2005\)](#). We choose the calibration period such that data for all countries in the considered group are available and the assumption of AR(1) around a linear trend is a reasonable assumption for the common latent effect. Results obtained with data on female mortality are discussed here; corresponding results for males are available in the online appendix.

Initializing prior distributions. We initialize the priors as discussed in the example with Swedish data (see Section 5.1).

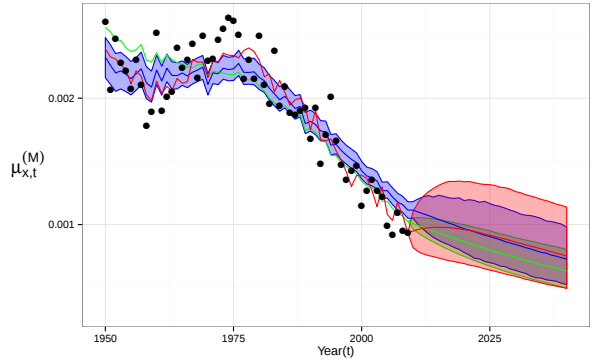
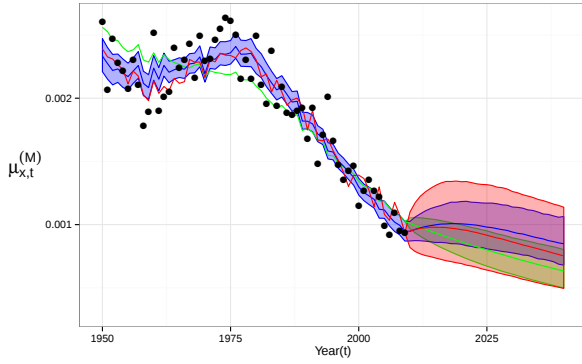
Estimates and credibility intervals in the LC-2,t model. Figure 8 shows the parameter estimates for $\alpha_x^{(i)}$ in the LC-2,t model and corresponding credibility intervals. We present results for Belgium, England-Wales and The Netherlands. These estimates are very close to the age specific intercepts obtained with a single population, country specific LC model. The intervals are narrow for all ages, similar to the findings of [Czado et al. \(2005\)](#).

Figures 9 and 10 show the estimates of $\beta_{x,1}^{(i)}$ and $\beta_{x,2}^{(i)}$ and their corresponding credibility intervals. For $\beta_{x,1}^{(i)}$ as well as $\beta_{x,2}^{(i)}$ the estimates corresponding to ages in the range from 0 to 35

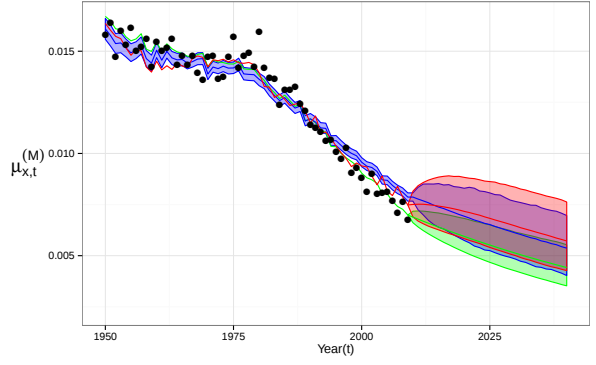
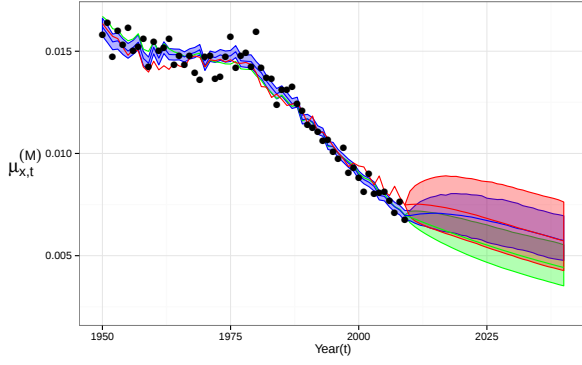
¹⁹See www.ag-ai.nl.



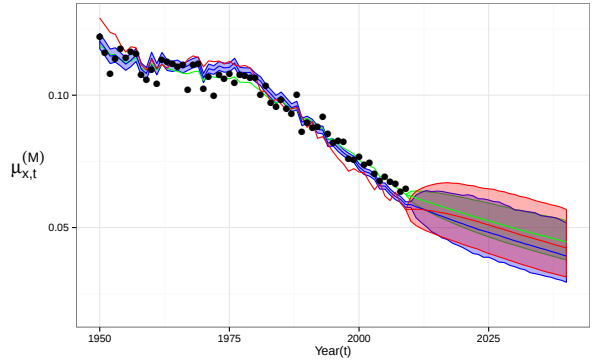
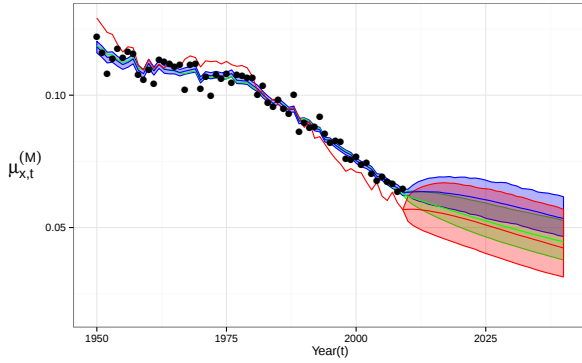
(a) Death rate fan charts ($\mu_{x,t}^{(M)}$) for Swedish males, LC-2,t (left) and LL (right) at age 20.



(b) Death rate fan charts ($\mu_{x,t}^{(M)}$) for Swedish males, LC-2,t (left) and LL (right) at age 40.

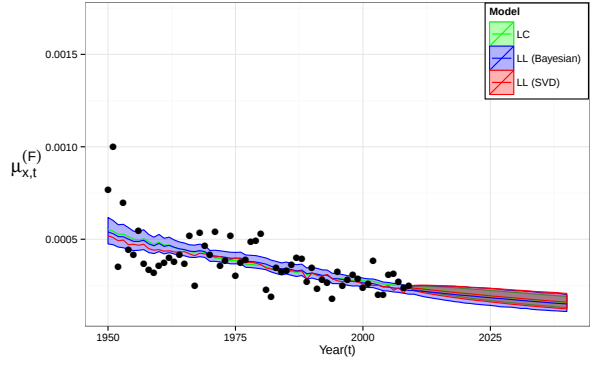
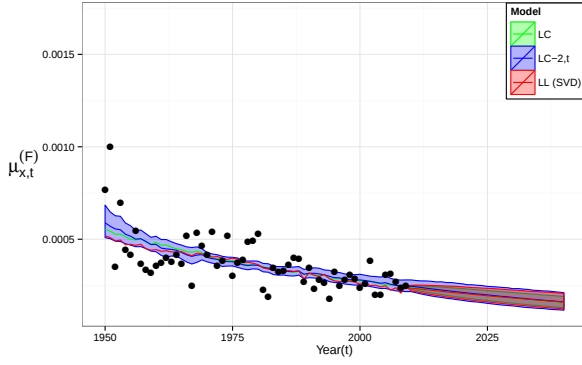


(c) Death rate fan charts ($\mu_{x,t}^{(M)}$) for Swedish females, LC-2,t (left) and LL (right) at age 60.

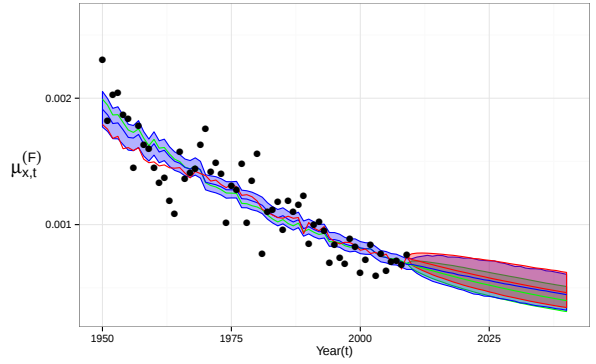
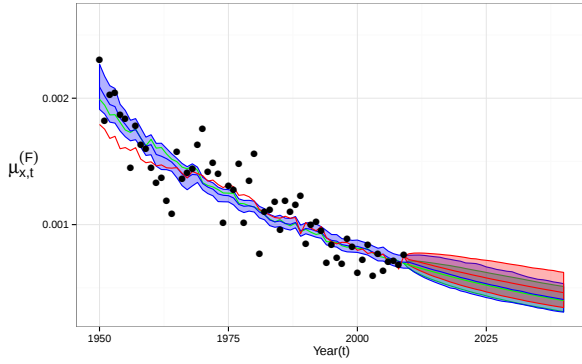


(d) Death rate fan charts ($\mu_{x,t}^{(M)}$) for Swedish males, LC-2,t (left) and LL (right) at age 80.

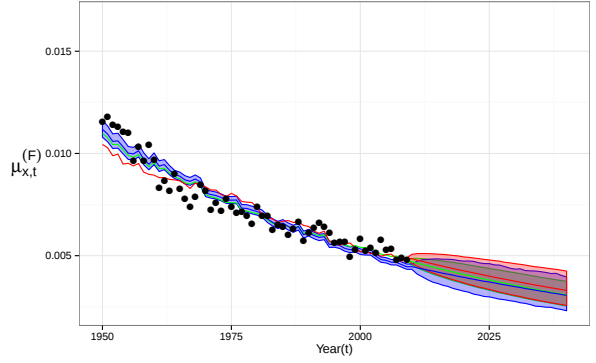
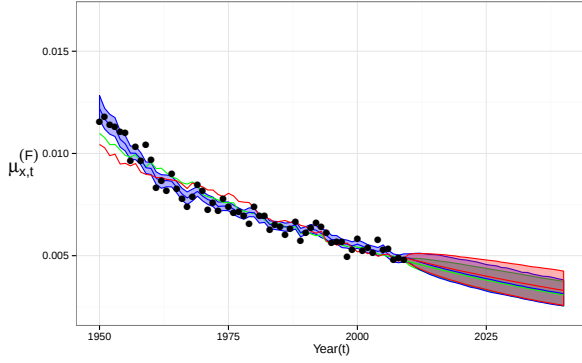
Figure 6: Death rate fan charts for Swedish males at age 20, 40, 60 and 80 for the LC,2-t (left) and LL (right) models.



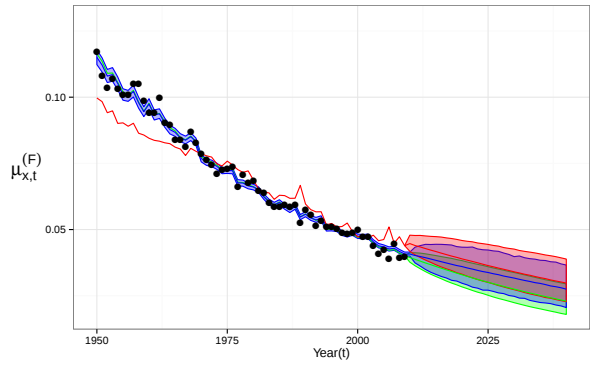
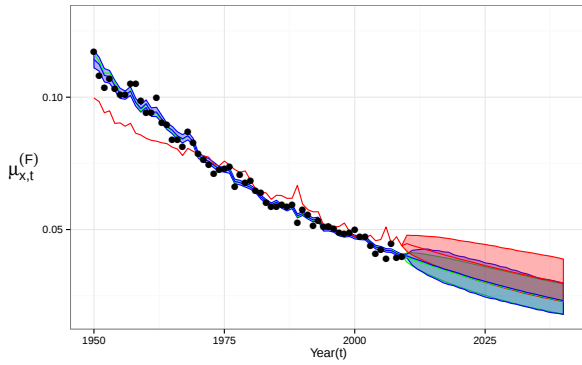
(a) Death rate fan charts ($\mu_{x,t}^{(F)}$) for Swedish males, LC-2,t (left) and LL (right) at age 20.



(b) Death rate fan charts ($\mu_{x,t}^{(F)}$) for Swedish males, LC-2,t (left) and LL (right) at age 40.



(c) Death rate fan charts ($\mu_{x,t}^{(F)}$) for Swedish females, LC-2,t (left) and LL (right) at age 80.



(d) Death rate fan charts ($\mu_{x,t}^{(F)}$) for Swedish females, LC-2,t (left) and LL (right) at age 80.

Figure 7: Death rate fan charts for Swedish females at age 20, 40, 60 and 80 for the LC,2-t (left) and LL (right) models.

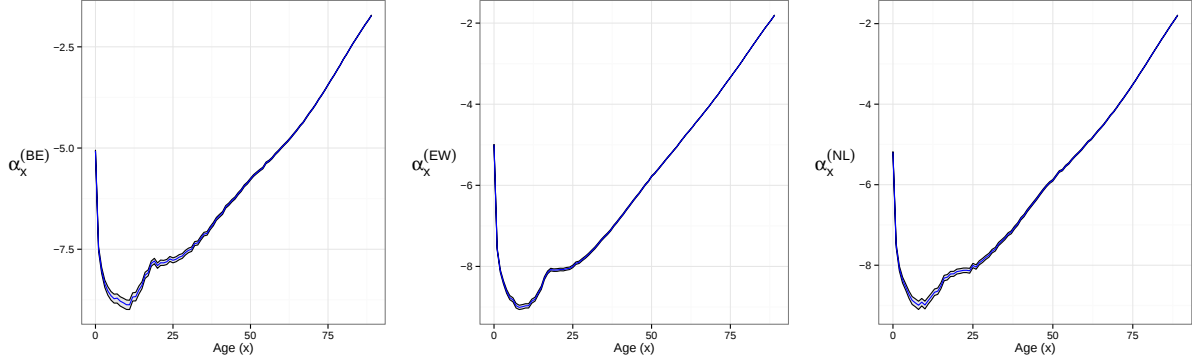


Figure 8: Model LC-2,t: parameter $\alpha_x^{(i)}$ for Belgium, England-Wales and The Netherlands: female data, period 1975-2009 and ages 0-89. 20,000 iterations, burn-in of 4,000 iterations and thinning factor 10. The blue line represents the median of the 1,600 iterations kept after thinning and burn-in. The shaded gray area is the 95% pointwise credibility interval using these iterations.

(approximately) show the largest uncertainty. Estimates corresponding to different countries show similarities, though differences are clearly visible.

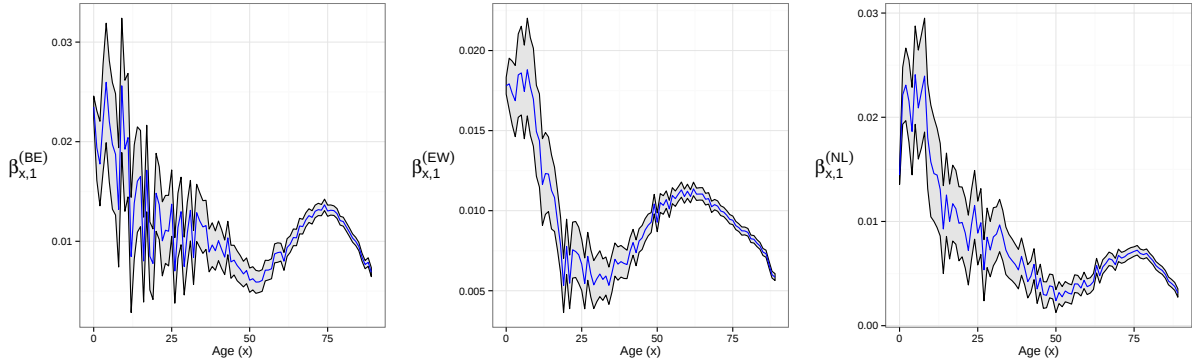


Figure 9: Model LC-2,t: parameter $\beta_{x,1}^{(i)}$ for Belgium, England-Wales and The Netherlands: female data, period 1975-2009 and ages 0-89. 20,000 iterations, burn-in of 4,000 iterations and thinning factor 10. The blue line represents the median of the 1,600 iterations kept after thinning and burn-in. The shaded gray area is the 95% pointwise credibility interval using these iterations.

Estimates and credibility intervals for the common latent effect, K_t , and the country specific latent effects, $\kappa_t^{(i)}$, are in Figure 11. Again we add the time trend of a LC model on $D_{x,t}^{(\bullet)}$ and $E_{x,t}^{(\bullet)}$, as well as the common trend obtained with SVD in the LL model, as reference lines. The posterior median of the intercept used in the linear trend prior of K_t is equal to 35.68 and the posterior median of the slope is -1.99 . The posterior median of ρ is 0.96. Country specific latent effects $\kappa_t^{(i)}$ are shown in the right panel of Figure 11 for a selection of countries (namely Belgium, England-Wales and The Netherlands). These country specific effects show mean reversion around 0, as specified in their time series prior, with posterior median of $\rho^{(BE)} = 0.79$, $\rho^{(EW)} = 0.76$ and $\rho^{(NL)} = 0.722$. The uncertainty in the $\kappa_t^{(i)}$ estimates of Belgium and the Netherlands is larger than the corresponding uncertainty for England-Wales.

Estimates and credibility intervals in the LL model. Figure 12 shows the parameter estimates of A_x , $\alpha_x^{(i)}$ and their corresponding credibility intervals. According to the intervals for

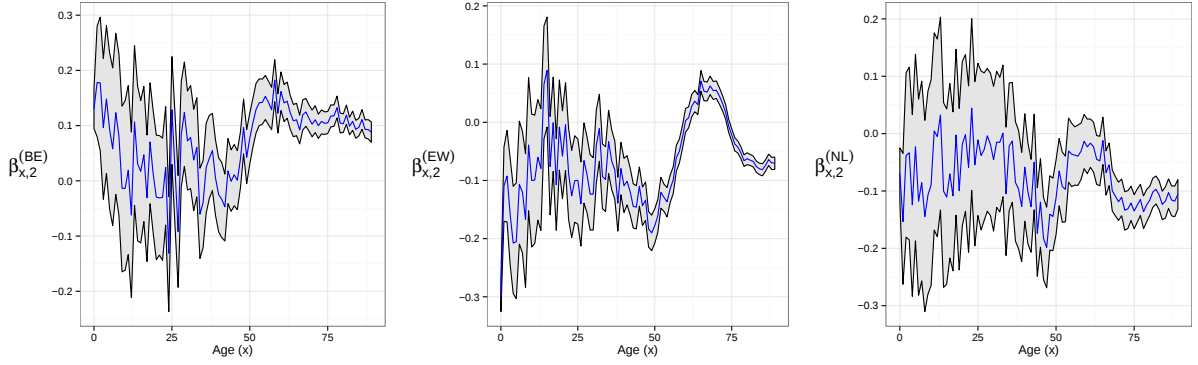


Figure 10: Model LC-2,t: parameter $\beta_{x,2}^{(i)}$ for Belgium, England-Wales and The Netherlands: female data, period 1975-2009 and ages 0-89. 20,000 iterations, burn-in of 4,000 iterations and thinning factor 10. The blue line represents the median of the 1,600 iterations kept after thinning and burn-in. The shaded gray area is the 95% pointwise credibility interval using these iterations.

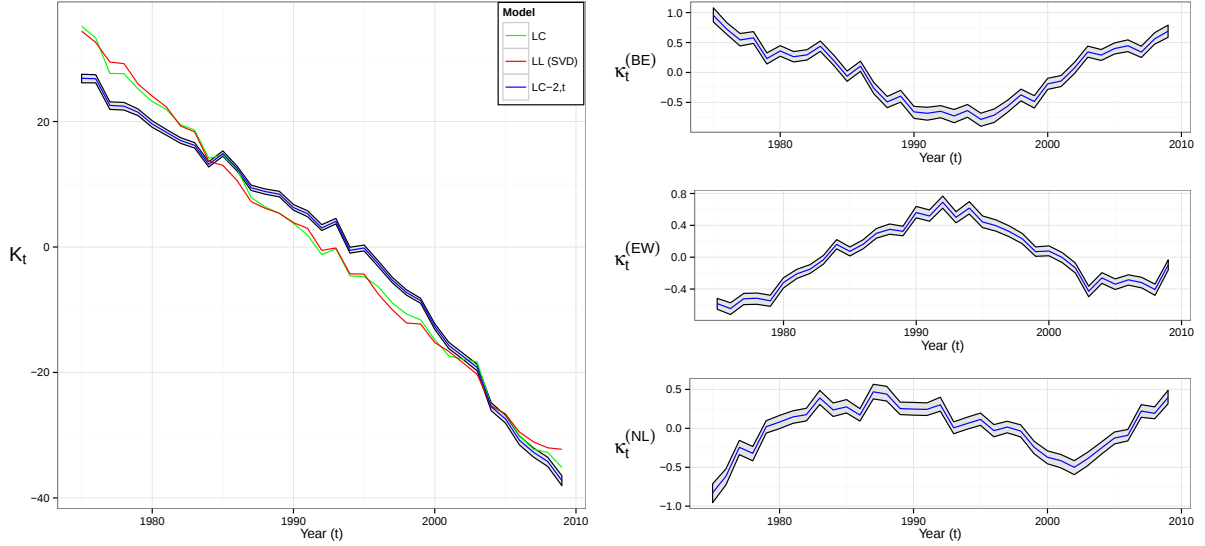


Figure 11: Model LC-2,t: parameters K_t and $\kappa_t^{(i)}$ for Belgium, England-Wales and The Netherlands: female data, period 1975-2009 and ages 0-89. 20,000 iterations, burn-in of 4,000 iterations and thinning factor 10. The blue line represents the median of the 1,600 iterations kept after thinning and burn-in. The shaded gray area is the 95% pointwise credibility interval using these iterations. The green reference line is $\hat{K}_t^{LC}$ and the red reference line is the common latent factor, obtained with SVD in the LL model.

the population specific $\alpha_x^{(i)}$ estimates corresponding to young ages are most uncertain.

Figure 13 shows the estimates of B_x and $\beta_x^{(i)}$ and their corresponding credibility intervals. The credibility intervals for the Belgian $\beta_x^{(i)}$ parameters enclose zero up to age 50.

Estimates and credibility intervals of the common latent effect, K_t , and the country specific latent effects, $\kappa_t^{(i)}$, are in Figure 14. The posterior median of the intercept in the linear trend prior for K_t is equal to 35.79 and the posterior median of its slope is -2.01 . The posterior median of $\rho = 0.953$. Country specific latent effects $\kappa_t^{(i)}$ for Belgium, England-Wales and The

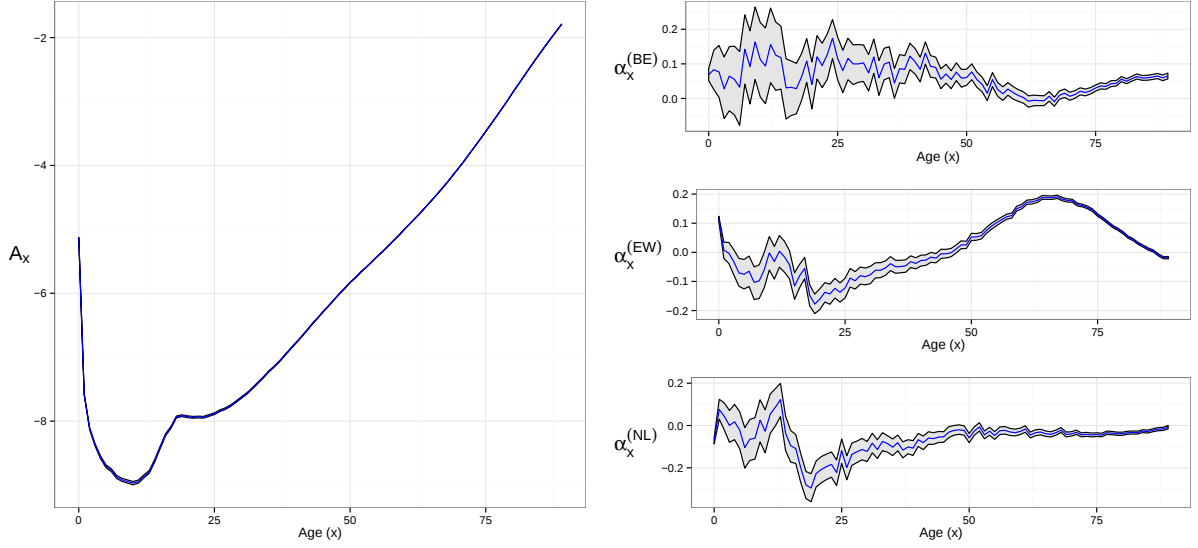


Figure 12: Model LL: parameters A_x and $\alpha_x^{(i)}$ for Belgium, England-Wales and The Netherlands: female data, period 1975-2009 and ages 0-89. 20,000 iterations, burn-in of 4,000 iterations and thinning factor 10. The blue line represents the median of the 1,600 iterations kept after thinning and burn-in. The shaded gray area is the 95% pointwise credibility interval using these iterations.

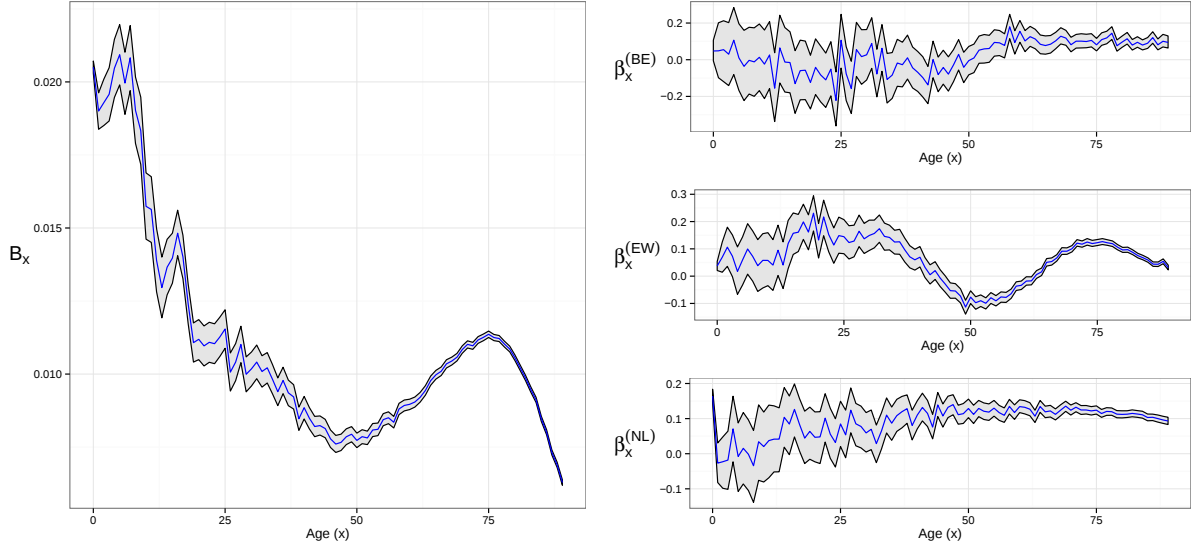


Figure 13: Model LL: parameters B_x and $\beta_x^{(i)}$ for Belgium, England-Wales and The Netherlands: female data, period 1975-2009 and ages 0-89. 20,000 iterations, burn-in of 4,000 iterations and thinning factor 10. The blue line represents the median of the 1,600 iterations kept after thinning and burn-in. The shaded gray area is the 95% pointwise credibility interval using these iterations.

Netherlands are shown in the right panel of Figure 11. These country specific effects have a posterior median of $\rho^{(\text{BE})} = 0.72$, $\rho^{(\text{EW})} = 0.83$ and $\rho^{(\text{NL})} = 0.88$. The uncertainty of $\kappa_t^{(i)}$ for Belgium and The Netherlands is larger than the corresponding uncertainty for England-Wales.

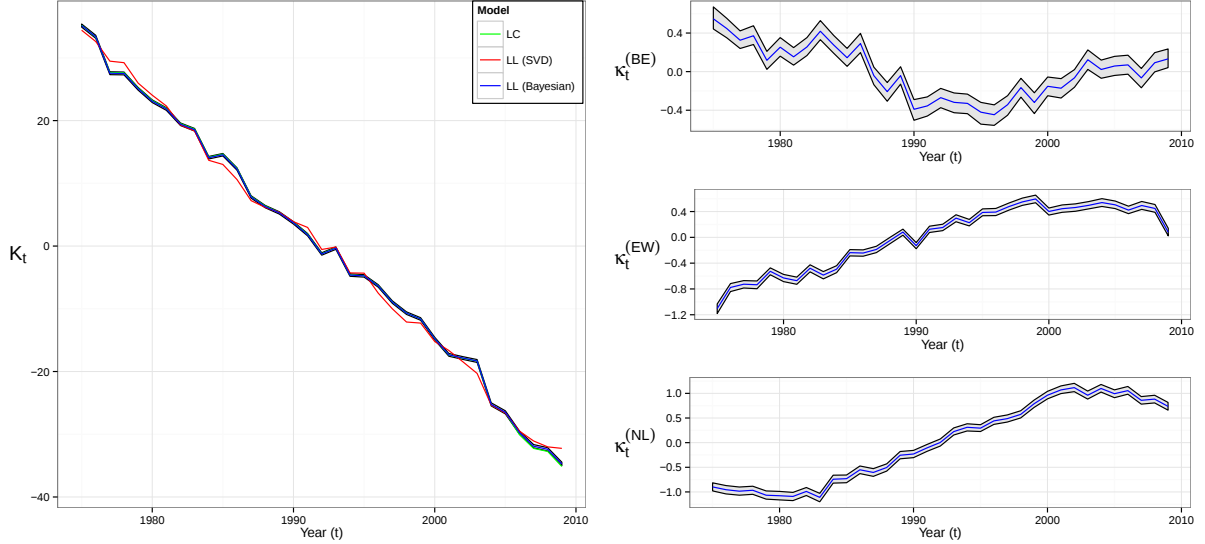
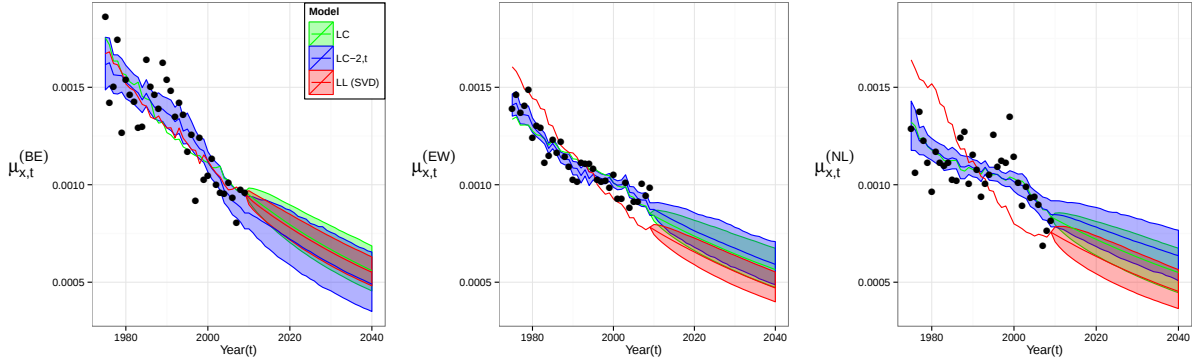


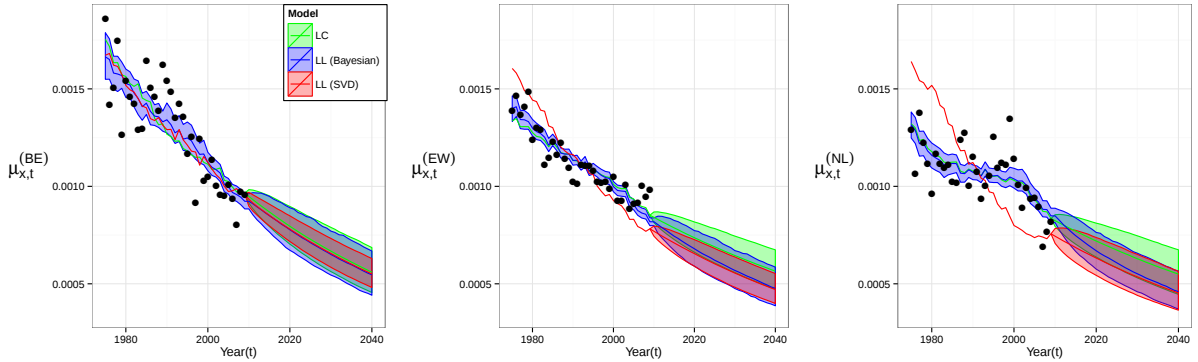
Figure 14: Model LL: parameters K_t and $\kappa_t^{(i)}$ for Belgium, England-Wales and The Netherlands: female data, period 1975-2009 and ages 0-89. 20,000 iterations, burn-in of 4,000 iterations and thinning factor 10. The blue line represents the median of the 1,600 iterations kept after thinning and burn-in. The shaded gray area is the 95% pointwise credibility interval using these iterations. The green reference line is $\hat{K}_t^{\text{LC}}$ and the red reference line is the common latent factor, obtained with SVD in the LL model.

Forecasting mortality: Belgium, England-Wales, The Netherlands. Figures 15 and 16 show calibrated death rates, $\mu_{x,t}^{(i)}$, their corresponding credibility intervals and forecasts for a selection of ages, namely the 40 and 60 years old for both Bayesian multi-population models. As in Section 5.1 we benchmark these results against the estimates obtained with a Poisson MLE implementation of a country specific LC model and the LL model calibrated using the Singular Value Decomposition. Both Bayesian multi-population models follow the observed data closely for all ages in the sample, whereas the SVD calibration of LL for some ages is not capturing well the observed evolution. This is clearly illustrated by the graphs of England-Wales and The Netherlands. The age 40 forecasts for Belgium and The Netherlands as obtained with the LC-2,t model are much more uncertain than the corresponding forecasts for England-Wales on the one hand and the forecasts obtained with the LL model on the other hand. The increase in uncertainty is due to the larger estimate and uncertainty of $\beta_{x,1}^{(i)}$ and $\beta_{x,2}^{(i)}$ in Figures 9 and 10.

Forecasting mortality: Austria versus West-Germany. A comparison between the death rate estimates and forecasts of a very large country, i.e. West-Germany, and a very small country, i.e. Austria, illustrates how uncertainty is reflected by the Bayesian multi-population models in the study. For a selection of ages Figure 17 and Figure 18 show the estimated and forecasted forces of mortality for Austria (top row) and West-Germany (bottom row). The results for Austria, with a population of approximately 8 million inhabitants, clearly reflect more uncertainty than the corresponding results for West-Germany.



(a) Model LC-2,t: Mortality fan charts for Belgium, England-Wales and The Netherlands at age 40.

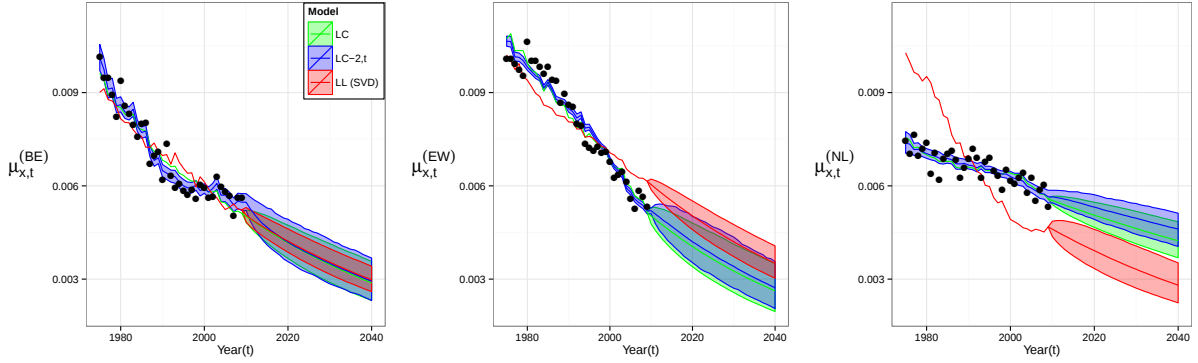


(b) Model LL: Mortality fan charts for Belgium, England-Wales and The Netherlands at age 40.

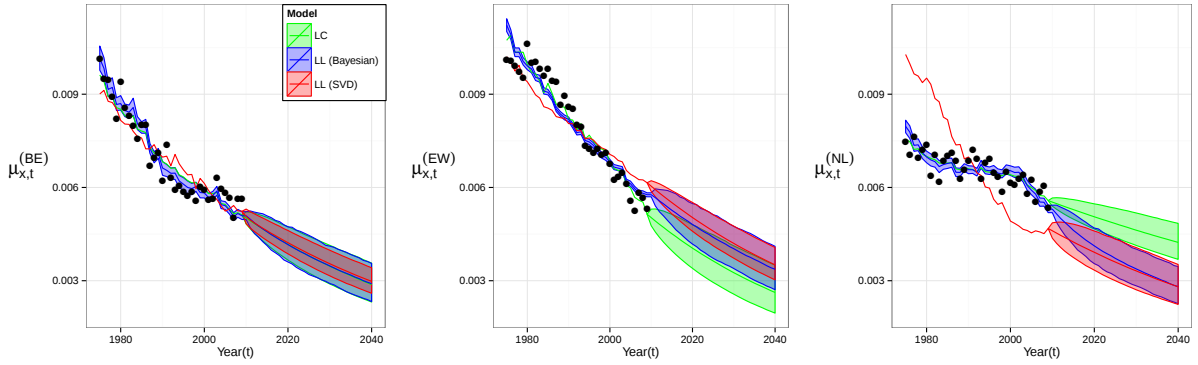
Figure 15: Mortality fan charts for Belgium, England-Wales and The Netherlands at age 40, using the LC-2,t model (top) and the LL model (bottom).

6 Conclusion

This paper presents a Bayesian implementation of two log-bilinear multi-population mortality models. The first model is the two factor version of the basic, single population Lee & Carter model, which we extend to the multi-population setting by replacing the first period effect with a common trend. This common period effect links the populations under study. The second model is the augmented common factor model, also known as the Li & Lee model. Multi-population mortality models recently have gained popularity at the expense of single population mortality models, as their use in official forecasts, hedging strategies and research illustrates. Multi-population mortality models use a larger data set to produce forecasts, allowing more robust mortality modeling by accounting for evolutions in mortality of related populations. We perform Bayesian estimation of both log-bilinear mortality models under the assumption of Poisson likelihood. We develop the complete technical details of the Markov Chain Monte Carlo updating scheme and extend as such the literature on Bayesian estimation of Poisson single population Lee & Carter type of models. Moreover this paper provides an alternative for the least squares calibration of log-bilinear multi-population mortality models on the one hand and the conditional Poisson maximum likelihood estimation on the other hand. Whereas maximum likelihood estimation requires a stepwise approach to mortality forecasting, i.e. estimate parameters in a first step and then project the estimated time dependent parameters in a second step, the use of a time series prior for the period effects enables a unified approach. Accounting for



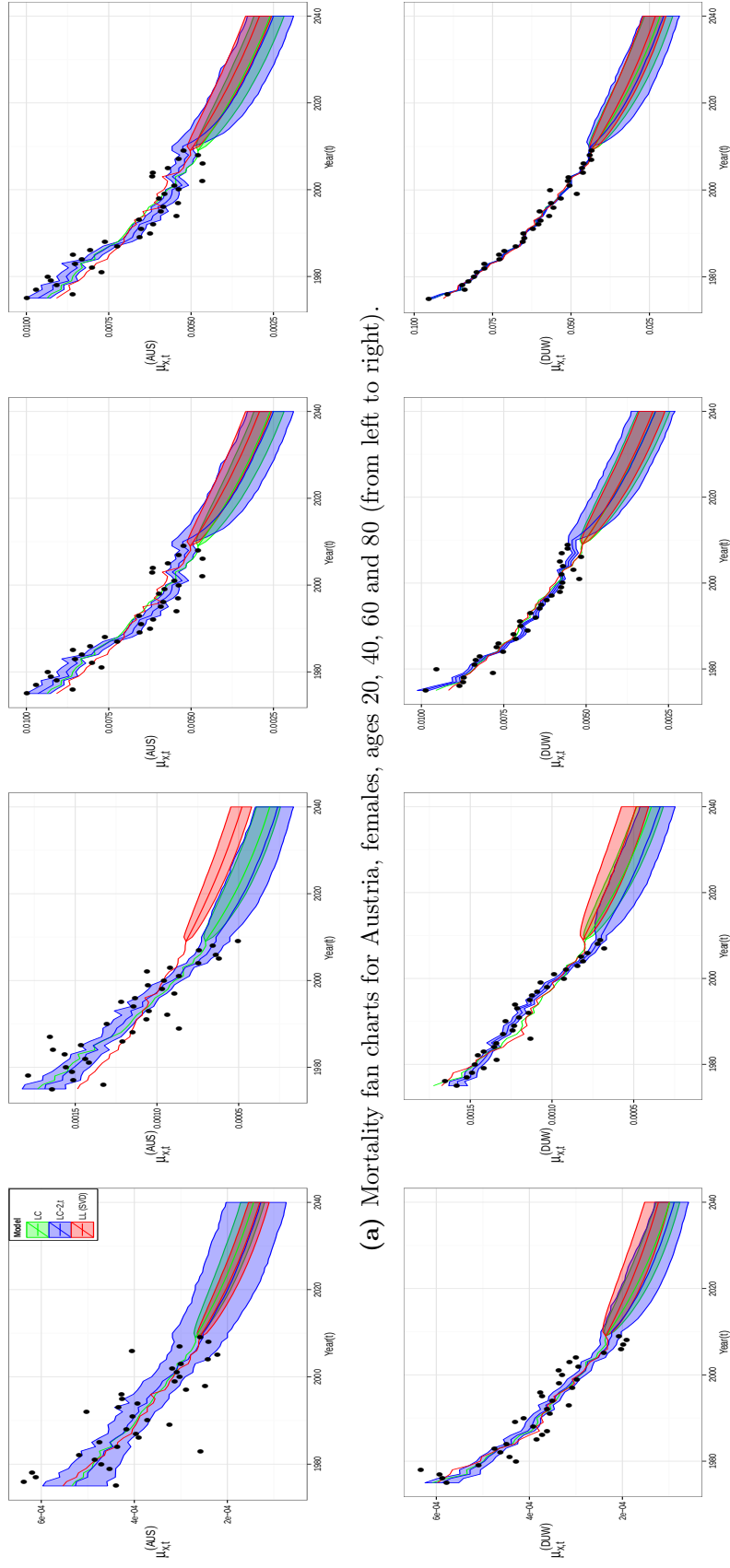
(a) Model LC-2,t: Mortality fan charts for Belgium, England-Wales and The Netherlands at age 60.



(b) Model LL: Mortality fan charts for Belgium, England-Wales and The Netherlands at age 60.

Figure 16: Mortality fan charts for Belgium, England-Wales and The Netherlands at age 60, using the LC-2,t model (top) and the LL model (bottom).

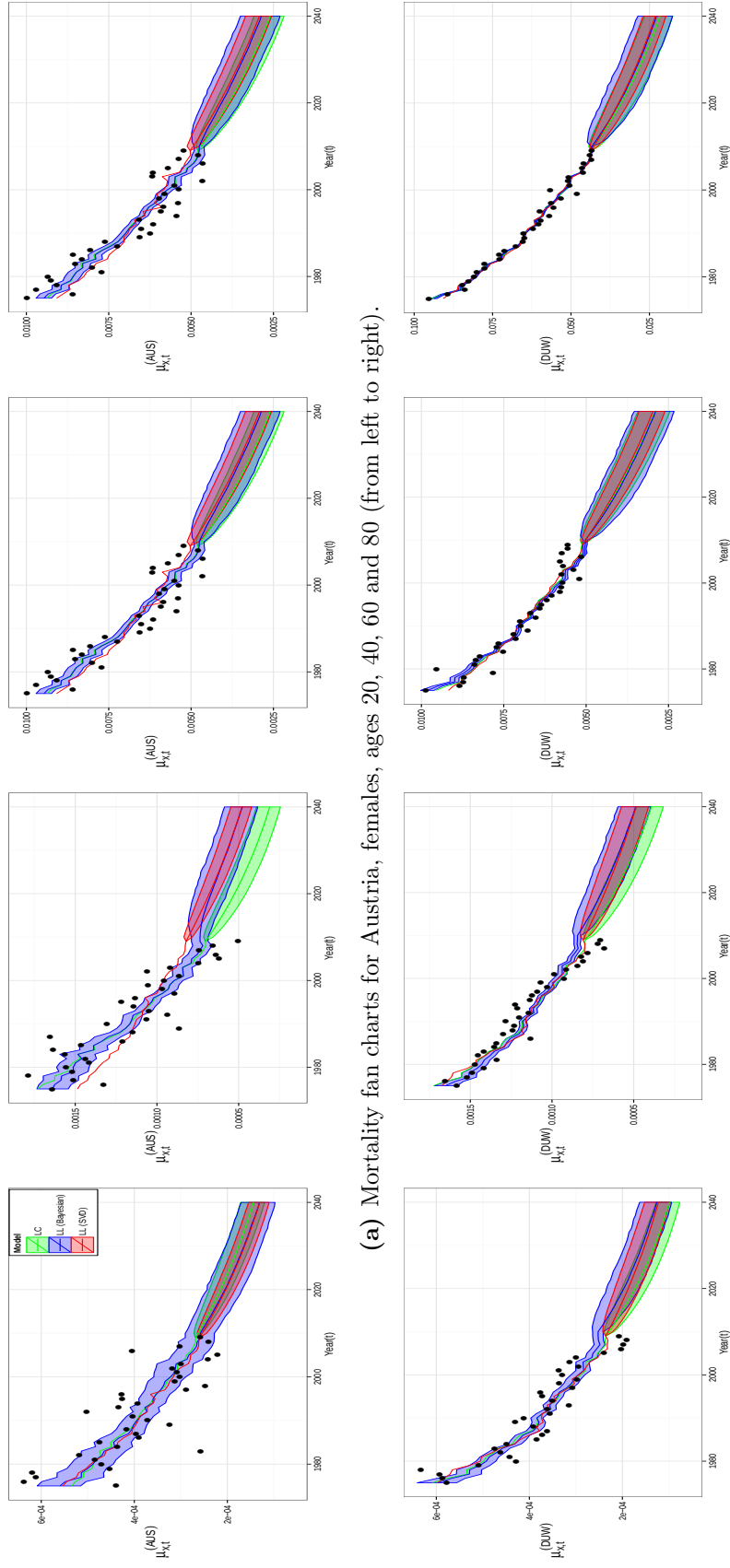
all sources of uncertainty is another direct advantage of the Bayesian approach and allows to present parameter estimates as well as their corresponding uncertainty. We demonstrate the use of the Bayesian log-bilinear multi-population models with two empirical studies, the analysis of Swedish male and female mortality during 1950-2009 and the analysis of a collection of 14 European countries during 1975-2009. The calibrated Poisson multi-population models follow the observed data closely, whereas the least squares estimation of the augmented common factor model for some calibration periods and specific ages does not capture well the observed mortality data. In future work we plan on developing the technicalities involved in the Bayesian estimation of other multi-population models. The extension of the present work to the joint modeling of portfolio and population mortality data is another direction for future research.



(a) Mortality fan charts for Austria, females, ages 20, 40, 60 and 80 (from left to right).

(b) Mortality fan charts for West-Germany, females, ages 20, 40, 60 and 80 (from left to right).

Figure 17: Model LC-2,t: mortality fan charts of a small (Austria, top row) versus a large (West-Germany, bottom row) country: females, ages 20, 40, 60 and 80 (from left to right), calibration period 1975-2009.



(a) Mortality fan charts for Austria, females, ages 20, 40, 60 and 80 (from left to right).

(b) Mortality fan charts for West-Germany, females, ages 20, 40, 60 and 80 (from left to right).
 Figure 18: Mortality fan charts of a small (Austria, top row) versus a large (West-Germany, bottom row) country: females, ages 20, 40, 60 and 80 (from left to right), calibration period 1975–2009.

A Metropolis-Hastings algorithm

A.1 Algorithm for K_t and $\kappa_t^{(i)}$

We use Metropolis-Hastings to update K_t . Assume that we have $K_s^{[l]}$ for all $s < t$ at iteration $[l]$. Then we proceed as follows:

1. Generate a candidate K_t^* from a normal distribution with mean $K_t^{[l]}$ and variance $s_{K_t}^2$. We start with $s_{K_t}^2 = 1$ and we update this value every 100 simulations to ensure that the acceptance probability is in the interval [20%, 50%]. For more details see [Czado et al. \(2005\)](#).
2. We use (37) and (38) to calculate the acceptance probability as follows:

$$\psi(K_t^{[l]}, K_t^*) = \min \left(1, \frac{f(K_t^* | \mathbf{K}_{-t}^{[l]}, \boldsymbol{\kappa}, \boldsymbol{\alpha}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{D}, \sigma_K^2, \rho, \gamma)}{f(K_t^{[l]} | \mathbf{K}_{-t}^{[l]}, \boldsymbol{\kappa}, \boldsymbol{\alpha}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{D}, \sigma_K^2, \rho, \gamma)} \right) \quad (47)$$

3. Generate a realization u from the Uniform(0, 1) distribution. If $u \leq \psi(K_t^{[l]}, K_t^*)$ we accept the candidate, i.e. $K_t^{[l+1]} = K_t^*$. If $u > \psi(K_t^{[l]}, K_t^*)$ we reject the candidate, so $K_t^{[l+1]} = K_t^{[l]}$.
4. Apply the constraint $\sum_{t=t_{min}}^{t_{max}} K_t = 0$.

The MH steps are similar for the parameters $\kappa_t^{(i)}$.

A.2 Algorithm for $\beta_{x,1}^{(i)}$ and $\beta_{x,2}^{(i)}$

Assume that we have $\beta_{x,1}^{(i)[l]}$ at iteration $[l]$. The Metropolis-Hastings updating scheme for $\beta_{x,1}^{(i)}$ goes as follows:

1. Generate a candidate $\beta_{x,1}^{(i)*}$ from a normal distribution with mean $\beta_{x,1}^{(i)[l]}$ and variance equal to $s_{\beta_1}^2 = 1$. The choice of the interval stems from empirical results. We start with variance $s_{\beta_1}^2$ and we update every 100 simulations to ensure that the acceptance probability is in the interval [20%, 50%]. For more details see [Czado et al. \(2005\)](#).
2. Calculate the acceptance probability as follows,

$$\psi(\beta_{x,1}^{(i)[l]}, \beta_{x,1}^{(i)*}) = \min \left\{ 1, \frac{f(\beta_{x,1}^{(i)*} | \boldsymbol{\beta}_{-x,1}^{(i)[l]}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \mathbf{K}, \sigma_{\beta_1}^2)}{f(\beta_{x,1}^{(i)[l]} | \boldsymbol{\beta}_{-x,1}^{(i)[l]}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \mathbf{K}, \sigma_{\beta_1}^2)} \right\} \quad (48)$$

3. Generate a realization u from the Uniform(0, 1) distribution. If $u \leq \psi(\beta_{x,1}^{(i)[l]}, \beta_{x,1}^{(i)*})$ we accept the candidate, i.e. $\beta_{x,1}^{(i)[l+1]} = \beta_{x,1}^{(i)*}$. If $u > \psi(\beta_{x,1}^{(i)[l]}, \beta_{x,1}^{(i)*})$ we reject the candidate, so $\beta_{x,1}^{(i)[l+1]} = \beta_{x,1}^{(i)[l]}$.
4. Apply constraint $\frac{1}{p} \sum_{i=1}^p \sum_{x=x_{min}}^{x_{max}} \beta_{x,1}^{(i)} = 1$.

The MH steps are similar for the parameters $\beta_{x,2}^{(i)}$.

A.3 Algorithm for ρ and $\rho_{(i)}$

Assume $\rho^{[l]}$ at iteration $[l]$ is available, then the MH updating scheme for ρ goes as follows.

1. Generate a candidate logit (ρ^*) from a normal distribution with mean logit ($\rho^{[l]}$) and variance $s_\rho^2 = 0.25^2$. This distribution is no longer symmetric. We start with variance s_ρ^2 and we update every 100 simulations to ensure that the acceptance probability is in the interval [20%, 50%]. For more details see [Czado et al. \(2005\)](#).
2. Calculate the acceptance probability as follows

$$\psi(\rho^{[l]}, \rho^*) = \min \left(1, \frac{f(\rho^* | \mathbf{K}, \gamma, \sigma_K^2)}{f(\rho^{[l]} | \mathbf{K}, \gamma, \sigma_K^2)} \frac{g(\rho^{[l]} | \rho^*)}{g(\rho^* | \rho^{[l]})} \right), \quad (49)$$

where g is the proposal distribution, namely the logit normal distribution and

$$\frac{g(\rho^{[l]} | \rho^*)}{g(\rho^* | \rho^{[l]})} = \frac{\rho^{[l]} (1 - \rho^{[l]})}{\rho^* (1 - \rho^*)}.$$

3. Generate a realization u from the Uniform(0, 1) distribution. If $u \leq \psi(\rho^{[l]}, \rho^*)$ we accept the candidate, $\rho^{[l+1]} = \rho^*$. If $u > \psi(\rho^{[l]}, \rho^*)$ we reject the candidate, so $\rho^{[l+1]} = \rho^{[l]}$.

The MH steps are similar for the hyperparameters $\rho_{(i)}$.

References

- Abel, G., Bijak, J., Forster, J., Raymer, J., and Smith, P. (2010). What do Bayesian methods offer population forecasters? *ESRC Centre for Population Change, Working Paper, 6*: http://www.cpc.ac.uk/publications/cpc_working_papers/pdf/2010_WP6_What_do_Bayesian_Methods_Offer_Population_Forecasters_Abel_et_al.pdf.
- Barrieu, P., Bensusan, H., Karoui, N. E., Hillairet, C., Loisel, S., Ravanelli, C., and Salhi, Y. (2012). Understanding, modelling and managing longevity risk : key issues and main challenges. *Scandinavian Actuarial Journal*, 3:203–231.
- Bray, I. (2002). Application of Markov Chain Monte Carlo methods to projecting cancer incidence and mortality. *Applied Statistics*, 3:3–43.
- Brouhns, N., Denuit, M., and Vermunt, J. (2002). A Poisson log-bilinear regression approach to the construction of projected life tables. *Insurance: Mathematics and Economics*, 31:373–393.
- Cairns, A., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73:687–718.
- Cairns, A., Blake, D., Dowd, K., Coughlan, G., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, 13(1):1–35.
- Cairns, A., Blake, D., Dowd, K., Coughlan, G., and Khalaf-Allah, M. (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, 41:29–59.

- Carter, L. and Lee, R. (1992). Forecasting demographic components: modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting*, 8:393–411.
- Congdon, P. (2006). *Bayesian Statistical Modelling*. Wiley, England.
- Czado, C., Delwarde, A., and Denuit, M. (2005). Bayesian Poisson log–bilinear mortality projections. *Insurance: Mathematics and Economics*, 36:260–284.
- Danesi, I., Haberman, S., and Millossovich, P. (2015). Forecasting mortality in subpopulations using Lee-Carter type models: A comparison. *Insurance: Mathematics and Economics*, 62(0):151–161.
- Dellaportas, P., Smith, A., and Stavropoulos, P. (2001). Bayesian analysis of mortality data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(2):275–291.
- Denuit, M. and Goderniaux, A.-C. (2005). Closing and projecting life tables using log-linear models. *Bulletin of the Swiss Association of Actuaries*, 1:29–48.
- Dowd, K., Blake, D., Cairns, A., Coughlan, G., and Khalaf-Allah, M. (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal*, 15:334–356.
- Fushimi, T. and Kogure, A. (2014). A Bayesian approach to longevity derivative pricing under stochastic interest rates with a two-factor Lee-Carter model. *ARIA 2014 Annual Meeting*: http://www.aria.org/Annual_Meeting/2014/2014_Accepted_Papers/4C/Fushimi%20and%20Kogure.pdf.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian data analysis*. Chapman & Hall/CRC Texts in Statistical Science, New York.
- Giroi, F. and King, G. (2008). *Demographic Forecasting*. Princeton: Princeton University Press.
- Haberman, S. and Renshaw, A. (2011). A comparative study of parametric mortality models. *Insurance: Mathematics and Economics*, 48:35–55.
- Hatzopoulos, P. and Haberman, S. (2013). Common mortality modeling and coherent forecasts. an empirical analysis of worldwide mortality data. *Insurance: Mathematics and Economics*, 52:320–337.
- Hyndman, R., Booth, H., and Yasmien, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50(1):261–283.
- Hyndman, R. and Ullah, M. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational statistics and data analysis*, 51:4942–4956.
- Jarner, S. and Kryger, E. (2011). Modelling adult mortality in small populations: The SAINT model. *ASTIN Bulletin*, 41(2):377–418.
- Kogure, A. and Yoshiyuki, K. (2010). A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance: Mathematics and Economics*, 46:162–172.
- Koninklijk Actuarieel Genootschap (2014). Prognosetafel AG2014. www.ag-ai.nl.
- Lee, R. and Carter, L. (1992). Modeling and forecasting the time series of US mortality. *Journal of the American Statistical Association*, 87:659–671.

- Li, H., Waegenare, A. D., and Melenberg, B. (2015). The choice of sample size for mortality forecasting: A bayesian learning approach. *Insurance: Mathematics and Economics*, (in press), doi: <http://dx.doi.org/10.1016/j.insmatheco.2015.03.024>.
- Li, J. (2013). A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies: A Journal of Demography*, 67(1):111–126.
- Li, J. and Hardy, M. (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal*, 15(2):177–200.
- Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee–Carter method. *Demography*, 42(3):575–594.
- Mardia, K., Kent, J., and Bibby, J. (2008). *Multivariate analysis*. Academic Press, Amsterdam.
- Mavros, G., Cairns, A., Kleinow, T., and Streftaris, G. (2014). A parsimonious approach to stochastic mortality modelling with dependent residuals. *Working paper*: <http://www.macs.hw.ac.uk/~andrewc/papers/MavrosEtAl2014.pdf>.
- Niu, G. and Melenberg, B. (2014). Trends in mortality decrease and economic growth. *Demography*, 51(5):1755–1773.
- Njenga, C. and Sherris, M. (2011). Modeling mortality with a Bayesian vector autoregression. *Working Paper*: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1776532.
- Pedroza, C. (2006). A Bayesian forecasting model: predicting U.S. male mortality. *Biostatistics*, 7:530–550.
- Pitacco, E., Denuit, M., Haberman, S., and Olivieri, A. (2009). *Modeling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press, London.
- Plat, R. (2009). On stochastic mortality models. *Insurance: Mathematics and Economics*, 45:393–404.
- Raftery, A., Chunn, J., Gerland, P., and Sevcikova, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography*, 50:777–801.
- Reichmuth, W. and Sarferaz, S. (2008). Bayesian demographic modelling and forecasting: An application to US mortality. *SFB 649 Discussion Paper 2008-052*: <http://edoc.hu-berlin.de/series/sfb-649-papers/2008-52/PDF/52.pdf>.
- Renshaw, A. and Haberman, S. (2003). Lee–Carter mortality forecasting with age specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272.
- Renshaw, A. and Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38:556–570.
- Riebler, A. and Held, L. (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models. *Biostatistics*, 11:57–69.
- Riebler, A., Held, L., and Rue, H. (2012). Estimation and extrapolation of time trends in registry data: borrowing strength from related populations. *Annals of Applied Statistics*, 6(1):304–333.
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.

- Stoeldraijer, L., van Duin, C., and Janssen, F. (2013a). Bevolkingsprognose 2012-2060: model en veronderstellingen betreffende de sterfte. *Bevolkingstrends*, Juli:1–27.
- Stoeldraijer, L., van Duin, C., van Wissen, L., and Janssen, F. (2013b). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: the case of The Netherlands. *Demography*, 29(13):323–354.
- Van Berkum, F., Antonio, K., and Vellekoop, M. (2014). The impact of multiple structural changes on mortality predictions. *Scandinavian Actuarial Journal*, DOI 10.1080/03461238.2014.987807.
- Wan, C., Berschi, L., and Yang, Y. (2013). Coherent mortality forecasting for small populations: an application to Swiss mortality data. Working Paper: http://www.actuaries.org/lyon2013/papers/AFIR_Wan_Bertschi_Yang.pdf.